

数学类专业学习辅导丛书

Guidance Series for Mathematics Majors

概率论基础学习指导书

李贤平 陈子毅 编著

概率论是一门研究随机现象数量规律的学科，它起源于17世纪中叶；其后，在对伯努利概型的深入研究中，发现了两种形式的极限定理——大数定律和中心极限定理，奠定了概率论在数学中的理论地位……



Guidance Series for Mathematics Majors

数学类专业学习辅导丛书

概率论基础学习指导书

Gailülun Jichu Xuexi Zhidaoshu

李贤平　陈子毅　编著



内容提要

本书以章前引言、课文导读、章后小议的新颖形式对《概率论基础》第三版的结构与内容、要点和难点作出独到的点评。

对占总数三成的基本题先作简要的提示后给答案，利于基础训练；对其余题目作出完整的解答以助充实提高；题后的评注则指明该题的命题含义、解题要点以及习题与正文的关联。

三十篇教学札记涉及基础概率论的许多理论、应用与历史专题，是作者四十多年教学的积累，对使用其他教材的师生也有较高的参考价值。

图书在版编目(CIP)数据

概率论基础学习指导书/李贤平，陈子毅编著.—北京：高等教育出版社，2011.7

ISBN 978-7-04-032238-5

Ⅰ.①概… Ⅱ.①李…②陈… Ⅲ.①概率论-高等学校-教学参考资料 Ⅳ.①O211

中国版本图书馆CIP数据核字(2011)第121851号

策划编辑 李蕊　　责任编辑 杨帆　　封面设计 张志　　版式设计 王艳红
插图绘制 郝林　　责任校对 杨雪莲　　责任印制 尤静

出版发行	高等教育出版社	咨询电话	400-810-0598
社　　址	北京市西城区德外大街4号	网　　址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印　　刷	北京宏信印刷厂	网上订购	http://www.landraco.com
开　　本	850mm × 1168mm　1/32		http://www.landraco.com.cn
印　　张	13.75	版　　次	2011年7月第1版
字　　数	350千字	印　　次	2011年7月第1次印刷
购书热线	010-58581118	定　　价	24.20元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换

物 料 号　32238-00

前　　言

本书由三大部分组成: 对《概率论基础》(第三版) 正文的点评、对它的习题的解答与评注、三十篇与基础概率论教学内容紧密相关的教学札记。

正文的点评, 包括章前引言、课文导读和章后小议, 对课文的重点与难点、结构和内容等作出评述。一般写得相当简练, 让读者在学习的各个阶段都能反复参照, 着眼于点化, 不炒冷饭, 不加罗列, 目的在于帮助学生更上一层楼。

作者深知, 习题解答的公开有利有弊。经慎重考虑采用如下方式处理: 对占总数三成的基本训练题只作切要的提示再给出简单答案, 对其余的则给出详细的解答, 并对几乎所有题目都写了评注。这样任课教师能根据课程要求对学生安排基本训练, 学生也能根据自己的能力与爱好作出进一步的选择。评注则用来指明命题的含义, 解题的要点或技巧, 与课文的关联以及题目之间的联系。

三十篇教学札记涉及概率的理论、应用与历史, 是笔者四十多年教学的积累, 这次借机形成文字, 以教师为主要对象作交流探讨, 也为学生提供课外阅读资料, 希望有助于他们对本学科的理解与深入。

教学札记的题材基本上有三类: 第一类以"×××和概率论"为标题, 介绍概率论的兄弟学科, 这些学科用概率语言作为表达本学科基本概念的工具, 初看它们是概率论的应用, 事实上是这些学科提高了概率论在现代科学中的地位。这里只涉及最重要的几个, 计有遗传学、统计物理学、量子力学、投资学、数理金融学、统计学。第二类, 对在概率论发展史中始终有较大影响的几个历史问题作些梳理, 包括分赌注问题、赌徒输光问题、圣彼得堡悖论、贝叶斯公式、正态刻画、切比雪夫不等式、博雷尔正规数等, 利用这个机会还编写了"概率论发展历史年表"作为全书附录。第三类, 与课文内容关系密切, 包括少量预备知识与若干外延, 都围绕一个个

专题写出。这些教学札记，每篇可独立阅读，整体编号，分列于最相关的章之后，暗示只要读完该章甚至前几节就可阅读。这部分材料对使用其他教材的师生有同等价值。

总之，本书虽以概率论基础学习指导书的形式出现，但编写时一直关注使内容对一般学习概率统计的学生，教概率统计的教师以及关心概率统计理论与应用的学者都有参考价值。深知该书一定存在不少缺点与错误，作者诚恳地希望得到批评与帮助。

这是我与陈子毅副教授的第二次合作。子毅是我的学生，后又成为同事。本书的编写分工如下：正文点评和教学札记由我编写；习题部分由子毅全部再做一遍，写出初稿，半年间我们共同反复修改不下五遍，并由我写下评注。电子版本的所有文字工作全由子毅承担。我们的合作十分愉快而有成效。对于子毅的工作，我的感觉是“放心”；对于与子毅的合作，我的检讨是“本应从 80 年代开始”。

对关心和帮助本书出版的朋友们表示感谢!

李贤平

2011. 2. 7.

目　录

导言 ·· 1

第一章　事件与概率 ·· 3

章前引言 ·· 3

课文导读 ·· 3

§1.1　随机现象与统计规律性 ·· 3

§1.2　样本空间与事件 ·· 4

§1.3　古典概型 ·· 6

§1.4　几何概率 ·· 8

§1.5　概率空间 ·· 8

习题解答与评注 ·· 9

习题总评 ·· 42

章后小议 ·· 42

教学札记之一　浅谈集合的大小 ······································ 43

教学札记之二　排列组合辑要 ·· 46

教学札记之三　对称性与概率计算 ···································· 54

教学札记之四　一般加法公式及其推广 ································ 61

第二章　条件概率与统计独立性 ······································ 67

章前引言 ·· 67

课文导读 ·· 67

§2.1　条件概率, 全概率公式, 贝叶斯公式 ···························· 67

§2.2　事件独立性 ·· 68

第四章　数字特征与特征函数 …… 233
章前引言 …… 233
课文导读 …… 233
§ 4.1　数学期望 …… 233
§ 4.2　方差, 相关系数, 矩 …… 234
§ 4.3　熵与信息 …… 236
§ 4.4　母函数 …… 237
§ 4.5　特征函数 …… 237
§ 4.6　多元正态分布 …… 239
习题解答与评注 …… 240
习题总评 …… 292
章后小议 …… 292
教学札记之二十　圣彼得堡悖论与期望效用函数 …… 293
教学札记之二十一　现代投资理论和概率论 …… 303
教学札记之二十二　数理金融学和概率论 …… 311
教学札记之二十三　事件的独立性与相关性 …… 321
教学札记之二十四　母函数与分支过程 …… 323

第五章　极限定理 …… 327
章前引言 …… 327
课文导读 …… 327
§ 5.1　伯努利试验场合的极限定理 …… 327
§ 5.2　收敛性 …… 329
§ 5.3　独立同分布场合的极限定理 …… 330
§ 5.4　强大数定律 …… 331
§ 5.5　中心极限定理 …… 333
习题解答与评注 …… 333
习题总评 …… 385
章后小议 …… 385
教学札记之二十五　统计学和概率论 …… 387
教学札记之二十六　统计学三大分布的推导 …… 394

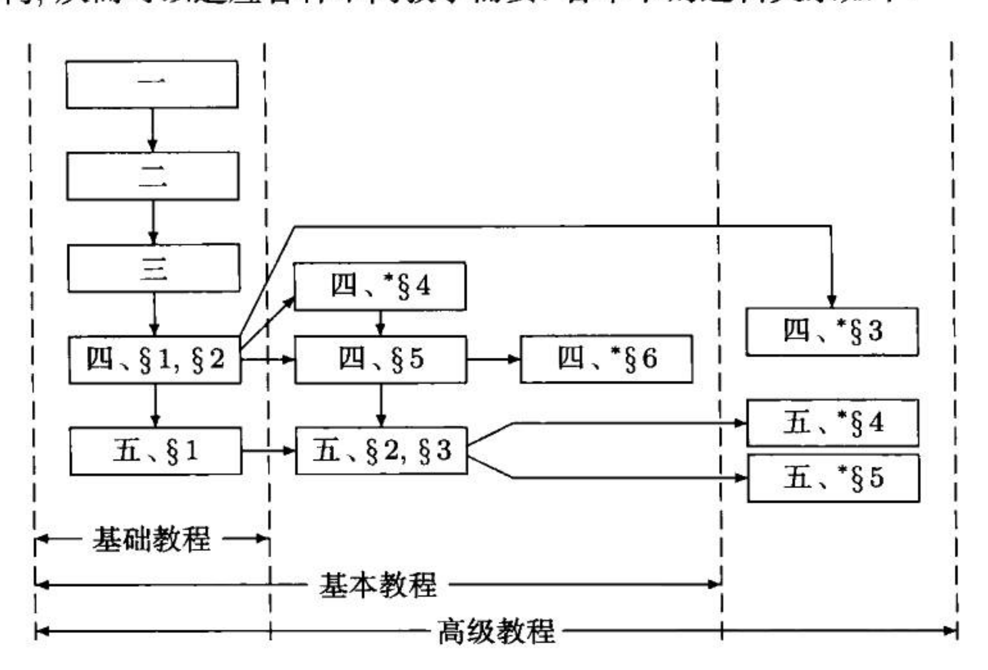
教学札记之二十七　[0, 1] 中数的 g 进制展开与概率论 ………… 401
教学札记之二十八　切比雪夫不等式与大数定律的证明 ………… 408
教学札记之二十九　强大数定律证明途径评述 ………… 415
教学札记之三十　关于进一步学习的建议 ………… 420

概率论发展历史年表 ………… 423

参考书目 ………… 428

导　言

《概率论基础》面向数学各专业和统计专业的学生, 只假定学生具有微积分基础知识, 因此也可供理工科、经济、管理甚至医农等学科中对概率论有较高要求的系科作为教材或教学参考书使用. 每周 4 学时, 可在一学期内讲完主体部分. 该书内容采用模块式结构, 从而可以适应各种不同教学需要. 各章节的逻辑关系如下:



因此《概率论基础》可在三种水平下使用, 主要区别在对极限定理的不同要求上.

第一种水平称为基础教程. 极限定理讲到伯努利试验场合, 即经典的伯努利大数定律与棣莫弗－拉普拉斯中心极限定理, 这两个定理可用较初等的方法证明.

第二种水平称为基本教程. 极限定理讲到独立同分布场合, 以特征函数为研究工具. 国内外的同类教科书大体就讲到这种水平, 已能满足数理统计等学科对概率论的要求.

第三种水平称为高级教程. 大数定律部分详细介绍以概率 1 收敛性和科尔莫戈罗夫强大数定律, 中心极限定理则讲到林德伯格–费勒定理.

总的说, 我国教材和教学大纲受苏联影响, 在概率论的极限定理部分讲得比欧美同类教材既广且深.

一般情况下, 讲授基本教程就够了, 但这主要应由任课教师根据具体情况自主决定.

第一章　事件与概率

章前引言

本章讲述了概率的严格定义.

机会游戏的对称性的利用和相对频率稳定性的观察在概率概念的建立中起重大作用.

机会游戏提供了研究不确定性的典型例子, 正是针对这种简单场合, 终于在 17 世纪中叶结晶出概率的概念, 从而宣布了概率论这一数学学科的诞生. 由于一批著名数学家的参与, 概率论的许多重要结果被建立, 也有了广泛的应用, 但是概率等基本概念的严格定义却经历了几乎整整三百年.

当代概率的定义是一种数学定义, 用公理化方式给出, 并未指明它的具体意涵. 本章结合概率概念的发展历史介绍这个过程, 详细讨论了古典概型、几何概率等具体模型, 导出许多基础性结果, 也为理解 (数学) 概率的实际涵义提供坚实的背景.

概率是对 (随机) 事件而言的, 因此事件与概率是本章的两个主角, 但是叙述还得从概率论的研究对象开始.

课文导读

§ 1.1　随机现象与统计规律性

本节作为导引, 介绍了概率论这一学科的研究对象与特色, 评

说了它的历史与前景, 围绕频率与概率的关系渐入主题, 为全书提供一个鸟瞰.

随机现象是广泛存在的, 因为有统计规律性即频率稳定性这一实践经验的存在, 使之值得研究也可以研究.

通过掷硬币、英文字母出现频率、高尔顿板等 3 个可以重复实验的权威性例子, 证明了频率稳定性是客观事实而非主观臆断. 师生们还可以进一步补充自己熟悉的例子.

频率与概率是概率论的永恒主题.

首先, 概率的频率解释还是当今最通行的解释, 在数理统计中频率学派仍占统治地位.

其次, 描述频率趋近于概率的大数定律总会是概率论的第一大定律.

第三, 在实际的工作中, 用频率来作为概率的估值是十分自然的.

《概率论基础》中提供的英文字母频率统计表是从大量出版物中统计出来的, 读者可试着任选一篇长度超过千个字符的英语文章作统计, 一般就能得出大致相似的结果. 因此利用字母出现频率就能破译早期那种用字母替换法编出的密码, 这可以设计成相当有趣的一个实验性游戏.

《概率论基础》第三版改以机票超售问题作为案例贯穿全书, 本节提出问题, 第二章设定概率模型, 最后, 在第五章给出解答.

关于概率论的历史, 参看本书最后的附录.

概率论的重要应用在《概率论基础》的后续章节陆续展开, 本书的教学札记中作了重要的补充.

§ 1.2 样本空间与事件

本节在集合论的基础上定义事件及其运算, 是以后一切讨论的基础, 务必牢固掌握.

事件与集合, 事件运算与集合运算的相似性, 这个事实也是长期研究后发现的, 目前已通行直接从这一事实出发, 定义事件及其

运算.

《概率论基础》认为对随机现象的研究联系到可以在相同条件下重复进行的“试验”, 而试验的可能出现的结果称为“样本点”, 样本点全体构成“样本空间” Ω. “事件”则定义为样本点的某个集合, 称某事件发生当且仅当它所包含的某一个样本点出现.

在这个定义下, 事件的关系及其运算可以与集合的关系及其运算建立如下对照表.

符号	概率论	集合论
Ω	样本空间, 必然事件	全集
$\varnothing$	不可能事件	空集
ω	样本点 (基本事件)	元素
A	事件	子集
$\overline{A}$	A 的对立事件 (逆事件)	A 的余集
$A\subset B$	事件 A 发生必有事件 B 发生	A 是 B的子集
$A=B$	事件 A 与事件 B 等价	A 与 B 相等
$A\cup B$	事件 A 与 B 至少有一个发生	A 与 B 的并集
$A\cap B$	事件 A 与 B 同时发生	A 与 B 的交集
$A-B$	事件 A 发生而事件 B 不发生	A 与 B之差
$AB=\varnothing$	事件 A 与事件 B 互不相容	A 与 B 没有公共元素

样本空间的概念来源于统计学, 它由统计学家冯・米泽斯 (von Mises, 1883—1953) 引入, 欧美教科书普遍采用. 我国的教材因受苏联教材影响, 过去较多称之为基本事件空间. 这个名称的缺点在于按《概率论基础》第一章 §5 给出的事件定义, 基本事件可以不是事件, 从而陷入窘境. 因此还是用样本空间为宜, 也便于与数理统计对接.

很快就会明白, 概率问题的复杂程度与所需使用的数学工具主要取决于样本空间的大小, 假如研究局限于离散样本空间, 则所用数学工具相当简单, 出了这个范围, 难度立刻大增. 为帮助没有系统学习过集合论的读者, 特撰写教学札记之一“浅谈集合的大小”,

扼要介绍有关集合大小的某些结果.

经验表明, 文图对理解事件的关系与运算帮助极大, 阅读和解题时都要充分利用它.

初学时要训练对某一事件能用集合论符号表出, 如《概率论基础》第一章 §2 例 7 所示, 接着要熟悉对某一集合论表达式能用概率论语言解释, 最后会感到二者实是一体.

关于事件的无限运算, 本节只引进可列并与可列交, 而不像某些教科书在这里就讲事件序列的极限, 以免吓退学生. 出于教学法的考虑, 为分散难点,《概率论基础》直至第五章 §4 才系统讨论事件序列极限.

集合运算相对于算术运算实在更为方便, 一是它成立对偶原理, 另一是它成立第二个分配律.

对偶原理是今后用得最多的恒等式, 在概率论中它把诸事件至少发生一个与它们全不发生这两个十分有用的事件联系在一起.

两事件等价, 如 $A = B$, 定义为 $A \subset B$ 及 $B \subset A$ 同时成立. 在论证中, 为了证明某两个事件等价, 基本途径还是要逐一验证这两个关系, 即从定义出发, 舍此之外别无良策. 学生们应踏踏实实做一两道这种题目, 以增加感性认识, 对偶原理及分配律都是很好的练习题. 当然随着学习的深入, 会推导出许多恒等式, 这些在今后学习中便可直接引用而不必再加以证明.

研究集合及其运算的另一方法 (数值化) 是引入示性函数, 这联系到以后要讲到的随机变量及数学期望等概念. 为使主线清晰, 本书未予强调, 但在第三章引入定义, 习题四 1 题中汇总.

§1.3 古典概型

具有有限与等可能性两大特征的古典概型是概率论的诞生地, 也是早期概率论研究的主要模型. 这时的概率是一个比例, 有相当直观的解释.

仔细思考, 会觉得人类要提炼出概率这么深奥的概念, 也只能

在这种最简单、最典型的场合. 它是各种机会游戏, 包括抛硬币、掷骰子、押宝、玩牌等等的模型化.

本节用近代的观点介绍了这个模型, 引导读者进入概率论.

古典概型的结果和研究方法被广泛吸收到自然科学 (特别是物理科学) 里, 表明现在仍未失去对古典概念和古典方法的兴趣. 关于在统计物理中的应用可参看教学札记之十五 “统计物理学和概率论”, 在遗传学中的应用参看教学札记之七 “遗传学和概率论”.

古典概率是一个比例, 是两个正整数之比, 为计算这个比例, 必须发展相应的计数方法. 事实上, 西方排列组合理论的发展与概率论的发展紧密相关. 教学札记之二 “排列组合辑要” 为读者提供一个关于排列组合的纲要: 其中 4 张表格对排列组合作各种分类; 例子则显示排列组合的基本应用方式, 大部分与课文相关; 长期与学生的接触中被问得最多的是关于重复组合公式, 故于该文利用不同题型给出 4 种不同证明.

用等可能性来定义概率 (事件发生的可能性), 逻辑上有些问题, 但在古典概型场合人们会觉得很自然, 主要是因为所考察的问题有一种对称性. 在解题中充分利用对称性, 可使某些问题化繁为简, 例如本节例 3. 教学札记之三 “对称性与概率计算” 提供更多例证.

摸球模型是很简单的一个模型, 也是很反映随机现象本质的一个模型, 因此在本书中反复被用来引进重要概念, 例如在本节引入了二项分布与超几何分布. 放回与不放回两种规则为何导致很不相同的结果, 这是每一个学习概率论的人必须搞清的, 其中的奥秘将随着本书的展开而逐步揭示. 教学札记之十三 “摸球与抽样” 提供某些参考资料.

古典概率计算大体分为直接法与公式法. 直接法只利用古典概率的定义, 以排列组合为工具计算有利场合数与样本点总数之比, 大体在本书已讲清. 公式法要用到概率的种种性质和推论, 将在以后不断补充.

鱼数估计提示一类主要统计问题和一种重要统计方法, 最早由拉普拉斯对人口统计提出.

古典概率三个基本性质为以后一般概率概念的定义作重要准备.

§ 1.4 几何概率

几何概率继续拉普拉斯的路线 —— 以等可能性定义概率. 在这种场合获得一定的成功, 也遇到新的困难.

几何概率处理无限的场合, 概率还是一个比例, 不过改成几何体的测度之比, 是两个正实数之比, 既是定义也是计算公式.

几何概率问题类型有限, 典型例子不多, 但蒲丰投针问题与蒙特卡罗积分计算有近代意义.

这些例子中隐含的假定将在第三章 §2 中挑明.

本节承前启后. 几何概率的成功与失败都对近代概率的定义有明显的促进.

本节首次出现事件的无限运算与概率的可列可加性.

§ 1.5 概率空间

水到渠成, 本节引入概率论的公理化结构 —— 概率空间 $(\Omega, \mathscr{F}, P)$.

主要的难点和要点在概率的定义域即事件域 $\mathscr{F}$ 的选择. 太小, 则事件太少, 不能满足需要; 太大, 则难以定义概率. 最后的折中选择是包含一切关心的事件的 σ 域, 后者保证了事件对交、并、逆、差作可列次运算的封闭性, 又把 $\varnothing$ 及 Ω 包含在内.

在这种 σ 域上, 能定义满足非负、规范和可列可加性的概率测度.

于是科尔莫戈罗夫把近代概率论奠基于测度论之上.

在离散样本空间中, $\mathscr{F}$ 可以取为 Ω 的一切子集, 而且只要把总和为 1 的概率分配给每个样本点就能完成整个概率空间的构造, 因此根本不用测度论. 这种场合构成了概率论的初等部分.

但是在样本空间为 $\mathbf{R}^1, \mathbf{R}^2, \mathbf{R}^n$ 等常用场合, 博雷尔 σ 域在严格定义概率论的各种概念中是不可缺少的. 《概率论基础》的方针是: 在满足概率论基本概念能严格定义的条件下, 尽量少用测度论. 本节是贯彻这个方针的起点.

概率论的公理化定义使概率的定义与概率的计算脱钩, 不像古典概型或几何概率那样. 概率的定义甚至与概率的解释脱离, 因此是一种数学概率论, 它容许对概率涵义的多种解释, 例如古典, 几何, 频率甚至主观的信念解释.

在以后的讨论中, 假定三位一体的概率空间 $(\Omega, \mathscr{F}, P)$ 已经给定.

从概率的三个公理能推导出概率的许多基本性质, 如欧几里得几何一样, 这些是我们已证得的定理, 在今后讨论中可以方便地使用. 导出的最复杂的性质是一般加法公式, 能用它解决经典名题"匹配问题". 进一步的讨论见教学札记之四"一般加法公式及其推广".

可列可加性等价于有限可加性与某种连续性 (上连续性, 下连续性, 甚至零上连续性), 从而保证了能进行极限运算. 应当指出这些讨论中只用到单调类: 即单调上升或单调下降的事件序列, 因此正如本章 §2 中指出的, 我们无须过早引入事件序列极限等较难的概念.

习题解答与评注

1. 在某城市中, 共发行三种报纸 A, B, C. 在这城市的居民中, 订阅 A 的占 45%, 订阅 B 的占 35%, 订阅 C 的占 30%, 同时订阅 A 及 B 的占 10%, 同时订阅 A 及 C 的占 8%, 同时订阅 B 及 C 的占 5%, 同时订阅 A, B, C 的占 3%, 试求下列百分率: (1) 只订阅 A 的; (2) 只订阅 A 及 B 的; (3) 只订阅一种报纸的; (4) 正好订阅两种报纸的; (5) 至少订阅一种报纸的; (6) 不订阅任何报纸的.

提示: 用文图, 细分析, 再求和.

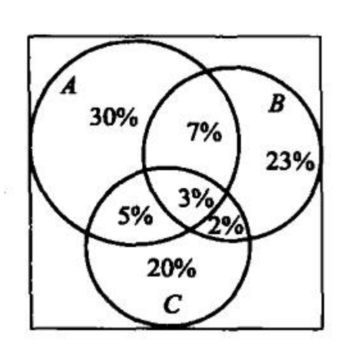


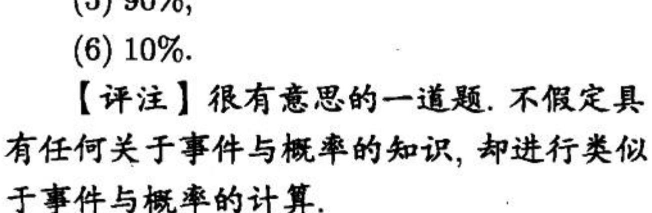
图 1-1

答 (1) 30%;

(2) 7%;

(3) 73%;

(4) 14%;



(5) 90%;

(6) 10%.

【评注】很有意思的一道题. 不假定具有任何关于事件与概率的知识, 却进行类似于事件与概率的计算.

A, B, C 把 Ω 分成不相交的 8 块, 至少需给定 7 个百分比才能算出有关答案.

本题可与《概率论基础》§2 例 7 连读.

2. 若 A, B, C 是随机事件, 说明下列关系式的概率意义: (1) $ABC = A$; (2) $A \cup B \cup C = A$; (3) $AB \subset C$; (4) $A \subset \overline{BC}$.

提示: 搞清集合论含义, 使用概率论语言 (字面表达上不强求统一).

答 (1) A 发生, 则 A, B, C 必同时发生; (2) B 发生或 C 发生时必有 A 发生; (3) A, B 同时发生必导致 C 的发生; (4) A 发生时, B 和 C 至少有一不发生.

【评注】熟识一下集合运算符号并能用概率论术语表述.

3. 在某班学生中任选一个同学, 以事件 A 表示选到的是男同学, 事件 B 表示选到的人不喜欢唱歌, 事件 C 表示选到的人是运动员. (1) 表述 $AB\overline{C}$ 及 $A\overline{B}C$; (2) 什么条件下成立 $ABC = A$; (3) 何时成立 $\overline{C} \subset B$; (4) 何时同时成立 $A = B$ 及 $\overline{A} = C$.

提示: 大胆用口语表述.

答 (1) $AB\overline{C}$={选到的学生是男同学, 又不爱唱歌, 又不是运动员}; $A\overline{B}C$={选到的是男同学, 又爱唱歌, 又是运动员}; (2) 当所有男同学都不爱唱歌且都是运动员时成立; (3) 当不是运动员的学

生必是不爱唱歌的学生时成立; (4) 男同学的全体就是不爱唱歌的学生全体, 也是非运动员的学生全体时成立.

【评注】具体事件的符号表达.

4. 试证:

$$\bigcup_{i=1}^{n} A_i = A_1 + \overline{A}_1 A_2 + \overline{A}_1\overline{A}_2 A_3 + \cdots + \overline{A}_1\overline{A}_2\overline{A}_3\cdots\overline{A}_{n-1}A_n,$$

并对 $n=4$, 画出文图.

证 设 $n=2$,

$$A_1 \bigcup A_2 = A_1 + (A_2 - A_1) = A_1 + \overline{A}_1 A_2$$

假设 $n=k$ 时成立

$$\bigcup_{i=1}^{k} A_i = A_1 + \overline{A}_1 A_2 + \overline{A}_1\overline{A}_2 A_3 + \cdots + \overline{A}_1\overline{A}_2\overline{A}_3\cdots\overline{A}_{k-1}A_k$$

则

$$\begin{aligned}\bigcup_{i=1}^{k+1} A_i &= \Big(\bigcup_{i=1}^{k} A_i\Big)\bigcup A_{k+1} \\ &= \bigcup_{i=1}^{k} A_i + \overline{\Big(\bigcup_{i=1}^{k} A_i\Big)} A_{k+1} \\ &= \bigcup_{i=1}^{k} A_i + \Big(\bigcap_{i=1}^{k} \overline{A}_i\Big) A_{k+1} \\ &= A_1 + \overline{A}_1 A_2 + \overline{A}_1\overline{A}_2 A_3 + \cdots + \\ &\quad \overline{A}_1\overline{A}_2\overline{A}_3\cdots\overline{A}_{k-1}A_k + \bigcap_{i=1}^{k}\overline{A}_i A_{k+1}\end{aligned}$$

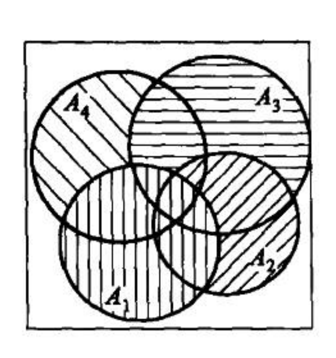


图 1-2

所以我们已用数学归纳法证得

$$\bigcup_{i=1}^{n} A_i = A_1 + \overline{A}_1 A_2 + \overline{A}_1\overline{A}_2 A_3 + \cdots + \overline{A}_1\overline{A}_2\overline{A}_3\cdots\overline{A}_{n-1}A_n$$

$n=4$ 时的文图见图 1-2.

【评注】把事件的并化为不相容事件之和的常见方式之一. 本题亦可用事件等价的定义证明.

若把 A_i 看作第 i 次时进入某禁地, 则上述表达式可称为"首次进入分解".

5. 为"剪刀·石头·布"游戏造一个样本空间, 定义有关事件, 并考虑如何给定概率.

答 样本空间 Ω={(剪刀, 剪刀), (剪刀, 石头), (剪刀, 布), (石头, 剪刀), (石头, 石头), (石头, 布), (布, 剪刀), (布, 石头), (布, 布)}.

以 (A_i, B_j) 表示甲、乙两人玩这游戏时, 甲出手形 A_i, 乙出手形 B_j, 其中 A_i, B_j, $i, j=1,2,3$, 分别表示剪刀, 石头, 布三种手形.

可定义事件{甲赢}={(剪刀, 布), (布, 石头), (石头, 剪刀)}, 类似可定义事件{乙赢}={(剪刀, 石头), (石头, 布), (布, 剪刀)}, 定义事件{和局}={(剪刀, 剪刀), (石头, 石头), (布, 布)}.

可认为样本空间中 9 个样本点是等可能出现的, 定义 $P\{(A_i, B_j)\}=\frac{1}{9}$, $i, j=1,2,3$.

于是, $P\{\text{甲赢}\}=P\{\text{乙赢}\}=P\{\text{和局}\}=\frac{1}{3}$.

【评注】对一方而言, 以等比例随机出剪刀, 石头, 布是最优策略. 双方都采用最优策略则出现上述等可能情况.

6. 若 A, B, C, D 是四个事件, 试用这四个事件表示下列各事件: (1) 这四个事件至少发生一个; (2) A, B 都发生而 C, D 都不发生; (3) 这四个事件恰好发生两个; (4) 这四个事件都不发生; (5) 这四个事件中至多发生一个.

提示: 《概率论基础》§2 例 7 的推广.

答 (1) $A\cup B\cup C\cup D$;

(2) $AB\overline{C}\,\overline{D}$;

(3) $AB\overline{C}\,\overline{D}+A\overline{B}C\overline{D}+\overline{A}BC\overline{D}+A\overline{B}\,\overline{C}D+\overline{A}B\overline{C}D+\overline{A}\,\overline{B}CD$;

(4) $\overline{A}\,\overline{B}\,\overline{C}\,\overline{D} = \overline{A \cup B \cup C \cup D}$;

(5) $\overline{A}\,\overline{B}\,\overline{C}\,\overline{D} + A\overline{B}\,\overline{C}\,\overline{D} + \overline{A}B\overline{C}\,\overline{D} + \overline{A}\,\overline{B}C\overline{D} + \overline{A}\,\overline{B}\,\overline{C}D$.

【评注】事件表示是学习概率论的入门功夫. 这种表达式不唯一, 一般喜欢写成: ①简洁形式; ②不相容事件之和; ③乘积形式.

*7. 从 0, 1, 2, $\cdots$, 9 中随机地取出 5 个数 (可重复), 以 E_i 记某些数正好出现 i 次这一事件 (例如 52353, 既属于 E_1, 也属于 E_2 及 E_0), 试用文图表示 E_0, E_1, $\cdots$, E_6 的关系.

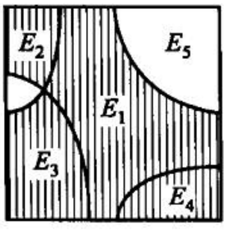


图 1－3

解 $E_0 = \Omega$, $E_6 = \varnothing$, $E_4 \subset E_1$,

$E_1E_5 = \varnothing$, $E_2E_5 = \varnothing$,

$E_3E_5 = \varnothing$, $E_4E_5 = \varnothing$,

$E_2E_4 = \varnothing$, $E_3E_4 = \varnothing$,

$E_1E_2E_3 = \varnothing$.

【评注】全部做对有点难度.

8. 证明下列等式:

(1) $\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n} = n2^{n-1}$

(2) $\binom{n}{1} - 2\binom{n}{2} + 3\binom{n}{3} - \cdots + (-1)^{n-1}n\binom{n}{n} = 0$

(3) $\sum\limits_{k=0}^{a-r}\binom{a}{k+r}\binom{b}{k} = \binom{a+b}{a-r}$

证 (1) 等式

$$1 + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n = (1+x)^n$$

两边对 x 求导, 得

$$\binom{n}{1} + 2\binom{n}{2}x + \cdots + n\binom{n}{n}x^{n-1} = n(1+x)^{n-1}$$

令 $x = 1$, 即得所要证的等式.

(2) 令上式中 $x = -1$ 即得.

(3) 展开等式

$$(1+x)^a(1+x)^b=(1+x)^{a+b}$$

$$\left(\sum_{m=0}^{a}\binom{a}{m}x^m\right)\left(\sum_{n=0}^{b}\binom{b}{n}x^n\right)=\sum_{l=0}^{a+b}\binom{a+b}{l}x^l$$

$$\sum_{l=0}^{a+b}\left(\sum_{j=0}^{l}\binom{a}{j}\binom{b}{l-j}\right)x^{j+l-j}=\sum_{l=0}^{a+b}\binom{a+b}{l}x^l$$

比较两边 x^l 前的系数, 就有

$$\sum_{j=0}^{l}\binom{a}{j}\binom{b}{l-j}=\binom{a+b}{l} \tag{*}$$

在 (*) 中取 $l=a-r\ (0\leqslant r\leqslant a)$, 则得

$$\sum_{j=0}^{a-r}\binom{a}{j}\binom{b}{a-r-j}=\binom{a+b}{a-r}$$

记 $a-r-j=k$, 则 $j=a-r-k$, 上式即为

$$\sum_{k=0}^{a-r}\binom{a}{a-r-k}\binom{b}{k}=\binom{a+b}{a-r}$$

也就是

$$\sum_{k=0}^{a-r}\binom{a}{k+r}\binom{b}{k}=\binom{a+b}{a-r}$$

【评注】用幂级数展开式证明组合恒等式是典型方法, 请读者自行补充新例. (*) 式即《概率论基础》的 (1.3.5), 是最重要的组合恒等式, 有明显的组合含义.

9. 一部五卷的文集, 按任意次序放到书架上去, 试求下列概率: (1) 第一卷出现在旁边; (2) 第一卷及第五卷出现在旁边; (3) 第一卷或第五卷出现在旁边; (4) 第一卷及第五卷都不出现在旁边; (5) 第三卷正好在正中.

提示: 作为《概率论基础》§3 例 1 的推广, 用全排列直接计算. 学过前两章之后再回过头来, 考虑其他解法.

答 (1) $\frac{2}{5}$; (2) $\frac{1}{10}$; (3) $\frac{7}{10}$; (4) $\frac{3}{10}$; (5) $\frac{1}{5}$.

【评注】古典概型常见类型之一: 全排列.

10. 甲袋中有 3 只白球, 7 只红球, 15 只黑球, 乙袋中有 10 只白球, 6 只红球, 9 只黑球, 现从两袋中各取一球, 求两球颜色相同的概率.

提示: 两袋各取一球配对构成样本点.

答 $\frac{207}{625}=0.331\,2$.

【评注】简单古典概率计算.

11. 袋中有 n 只球, 记有号码 $1,2,\cdots,n$, 求下列事件的概率: (1) 任意取出 2 球, 号码为 1, 2; (2) 任意取出 3 球, 没有号码 1; (3) 任意取出 5 球, 号码 1, 2, 3 中至少出现 1 个.

提示: 不计顺序, 宜用组合.

答 (1) $\frac{2}{n(n-1)}$; (2) $\frac{n-3}{n}$; (3) $1-\frac{\binom{n-3}{5}}{\binom{n}{5}}$.

【评注】古典概型常见类型之二: 不放回摸球.

12. 袋中装有 $1,2,\cdots,N$ 号的球各一只, 采用 (1) 有放回; (2) 不放回方式摸球, 试求在第 k 次摸球时首次摸到 1 号球的概率.

提示: (1) 用重复排列; (2) 用全排列或用对称性.

答 (1) $\frac{(N-1)^{k-1}}{N^k}$; (2) $\frac{1}{N}$.

【评注】重要的模型, 有名的答案—— 开锁即为一例. 两种不同摸球方式, 导致很不相同答案.

13. 从 6 双不同的手套中任取 4 只, 问其中恰有一双配对的概率是多少?

提示: 从 12 只中任取 4 只构成样本空间, 再细算有利场合.

答 $\frac{16}{33}$.

【评注】大约有近十种做法可得出这一正确答案.

14. 从 n 双不同的鞋子中任取 $2r$ $(2r < n)$只, 求下列事件发生的概率: (1) 没有成对的鞋子; (2) 只有一对鞋子; (3) 恰有两对鞋子; (4) 有 r 对鞋子.

解 (1) 有利场合是: 先从 n 双不同的鞋子中取出 $2r$ 双, 再从每双中任取 1 只, 概率为

$$\frac{\binom{n}{2r}2^{2r}}{\binom{2n}{2r}}$$

(2) 有利场合是: 先从 n 双中取出 1 双, 其 2 只全都取出, 再从余下的 $n-1$ 双中取出 $2r-2$ 双, 在每双中只取 1 只, 概率为

$$\frac{\binom{n}{1}\binom{n-1}{2r-2}2^{2r-2}}{\binom{2n}{2r}}$$

(3) 仿照 (2) 的思路, 得概率

$$\frac{\binom{n}{2}\binom{n-2}{2r-4}2^{2r-4}}{\binom{2n}{2r}}$$

(4) 有 r 对鞋子, 概率为

$$\frac{\binom{n}{r}}{\binom{2n}{2r}}$$

【评注】上题的一般化.

15. m 个男孩和 n 个女孩 $(n \leqslant m)$ 随机地沿着圆桌坐下, 试求任意两个女孩都不相邻的概率.

解 m 个男孩排成一直线, 有 $m!$ 种排列法. 这 m 个男孩如排成一圆圈, 则直线上 m 种排列法对应于圆圈上的一种排列法 (例如以顺时针计), 所以 m 个男孩沿圆桌坐下的圆形排列共有 $\frac{m!}{m}=(m-1)!$ 种. 沿圆桌坐的 m 个男孩间有 m 个空位来安插 n 个女孩, 能使任意两个女孩都不相邻的坐法有 $\binom{m}{n}n!$ 种. 样本点总数应为 $m+n$ 个孩子沿圆桌坐下的所有可能圆形排列数. 于是, 所求概率为

$$\frac{(m-1)!\binom{m}{n}n!}{(m+n-1)!}=\frac{\binom{m}{n}}{\binom{m+n-1}{n}}$$

【评注】以上解法用圆形排列得到答案, 也可用圆形组合直接给出右端答案.

16. (分赌注问题) 甲、乙二人各出赌注 a, 约定谁先胜三局则赢得全部赌注, 现已赌三局, 甲二胜一负, 这时因故终止赌博, 若二人赌技相同, 问应如何分配赌注, 才算公平合理?

提示: 依再赌下去两人得胜的概率来分配赌注.

答 $3:1$.

【评注】本题是教学札记之八 "分赌注问题" 中提到的导致帕斯卡与费马通信并催生概率论的问题的简化形式, 当时帕斯卡, 费马和惠更斯都用不同思路与方法得到上述正确答案. 希望你也能给出自己的答卷. 下章将对这个问题作一般性讨论.

17. 从 52 张扑克牌中任意取出 13 张, 求: (1) 有 5 张黑桃, 3 张红心, 3 张方块, 2 张草花的概率; (2) 牌型分布为 7-3-2-1 (最长花色有 7 张, 最短花色有 1 张, 其余二花色分别有 3 张及 2 张) 的概率.

提示: (1) 从 52 张牌中取 13 张构成样本空间, 不考虑顺序.

(2) 化为 (1).

答 (1) $\dfrac{\dbinom{13}{5}\dbinom{13}{3}\dbinom{13}{3}\dbinom{13}{2}}{\dbinom{52}{13}}$.

(2) $\dfrac{\dbinom{13}{7}\dbinom{13}{3}\dbinom{13}{2}\dbinom{13}{1}}{\dbinom{52}{13}} \times 4 \times 3 \times 2$.

【评注】桥牌中牌型分布是一手牌强弱的主要因素之一, 桥牌叫牌体系设计中牌型分布概率的计算至关重要. 桥牌在所有竞技项目中是使用概率论最多的.

18. 桥牌游戏中 (四人各从 52 张纸牌中分得 13 张), 求 4 张 A 集中在一个人手中的概率.

提示: 参照上题.

答 $\dfrac{\dbinom{4}{4}\dbinom{48}{9}}{\dbinom{52}{13}} \times 4$.

【评注】大牌是桥牌中一手牌强弱的另一个主要因素, 其分配方式也提出大量概率计算问题.

*19. 在扑克牌游戏中 (从 52 张牌中任取 5 张), 求下列事件的概率: (1) 以 A 打头的同花顺次五张牌; (2) 其他同花顺次五张牌; (3) 有四张牌同点数; (4) 三张同点数且另两张也同点数; (5) 五张同花; (6) 异花顺次五张牌; (7) 三张同点数, 另外两张不同点数; (8) 五张中有两对; (9) 五张中有一对; (10) 其他情况.

提示: 52 张牌中取 5 张不计序, 样本点总数为 $\dbinom{52}{5}$. 重点在计算有利场合数. 注意两点游戏约定: ①样本点若已归入前面的事件 (强的), 则不再归入后面的事件 (弱的). ②约定 A 可放在 K, Q,

J 之前, 也可放在 4, 3, 2 之后构成顺次.

答 所求概率如下:

分母为 $\binom{52}{5} = 2\,598\,960$.

分子分别为:

英文名称	分子	概率
(1) Royal flush	$\binom{4}{1}$	0.000 001 54
(2) Other straight flushes	$\binom{4}{1}\binom{9}{1}$	0.000 013 85
(3) Four of a kind	$\binom{13}{1}\binom{4}{4}\binom{48}{1}$	0.000 240
(4) Full house	$\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}$	0.001 441
(5) Flush	$\binom{4}{1}\binom{13}{5}-\binom{4}{1}-\binom{4}{1}\binom{9}{1}$	0.001 965
(6) Straight	$\binom{10}{1}\binom{4}{1}^5-\binom{10}{1}\binom{4}{1}$	0.003 925
(7) Three of a kind	$\binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}\binom{4}{1}$	0.021 128
(8) Two pairs	$\binom{13}{2}\binom{4}{2}\binom{4}{2}\binom{11}{1}\binom{4}{1}$	0.047 539
(9) One pair	$\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3$	0.422 569
(10) Other		$1-\sum_{i=1}^{9} p_i = 0.501\,177$

【评注】取材于一种在西方流传很广的随机游戏. 答案中的概率数值依次上升, 表明相应各组牌的强度依序下降.

以上三题虽以纸牌游戏形式给出, 实是古典概型的一种典型题, 属于超几何分布及其推广.

【劝告】以下几题是读者考验自己进行事件分析和古典概率计算能力的好题目, 务必经过自己反复思考再看答案, 以免糟蹋

题目.

20. 从装有号码 $1,2,\cdots,N$ 的球的箱子中有放回地摸了 n 次球, 依次记下其号码, 试求这些号码按严格上升次序排列的概率.

提示: 要害在"严格上升".

答 $\dfrac{1}{N^n}\dbinom{N}{n}$.

【评注】看了答案就平淡无奇.

21. 在上题中这些号码按上升 (不一定严格) 次序排列的概率.

答 如果这些号码不一定是严格上升的次序, 意味着 n 个号码可以重复, 这就相当于从 N 个不同元素中取 n 个元素的重复组合, 由此得本题所求的概率为 $\dfrac{1}{N^n}\dbinom{N+n-1}{n}$.

【评注】重复组合数的推导见教学札记之二"排列组合辑要".

22. 任意从数列 $1,2,\cdots,N$ 中不放回地取出 n 个数并按大小排列成: $x_1 < x_2 < \cdots < x_m < \cdots < x_n$, 试求 $x_m = M$ 的概率, 这里 $1 \leqslant M \leqslant N$.

提示: 考虑 $x_m = M$ 之下有几个小于 M, 几个大于 M.

答 $\dfrac{\dbinom{M-1}{m-1}\dbinom{1}{1}\dbinom{N-M}{n-m}}{\dbinom{N}{n}}$.

【评注】预告多元超几何分布, 见《概率论基础》(3.2.7).

*23. 上题中, 若采用有放回取数, 这时 $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_m \leqslant \cdots \leqslant x_n$, 试求 $x_m = M$ 的概率.

解 从 $1,2,\cdots,N$ 中有放回地取 n 个数, 这 n 个数可分为 3 类: 小于 M, 等于 M, 大于 M. 以 k_1 记取到小于 M 的数的次数, 以 k_2 记取到大于 M 的数的次数, 则取到等于 M 的次数为 $n-k_1-k_2$.

在固定 k_1, k_2 的条件下, 取到 k_1 个小于 M 的数, $n-k_1-k_2$

个 M, k_2 个大于 M 的数的概率为

$$\frac{n!}{k_1!k_2!(n-k_1-k_2)!}\frac{(M-1)^{k_1}(N-M)^{k_2}}{N^n}$$

易见 k_1 可取 $0,1,2,\cdots,m-1$, k_2 可取 $0,1,2,\cdots,n-m$, 于是所求概率为

$$\sum_{k_1=0}^{m-1}\sum_{k_2=0}^{n-m}\frac{n!}{k_1!k_2!(n-k_1-k_2)!}\frac{(M-1)^{k_1}(N-M)^{k_2}}{N^n}$$

【评注】事件分析对复杂概率计算十分重要, 本题是一个很好的例子. 本题亦预告多项分布, 见《概率论基础》(3.2.6).

24. 从 (0,1) 中随机地取两个数, 求下列概率: (1) 两数之和小于 1.2; (2) 两数之积小于 0.25; (3) 以上两个要求同时满足.

解 从 (0,1) 中取出两数分别记为 x, y, 则 (x,y) 与单位正方形 $ABCD$ 内的点一一对应.

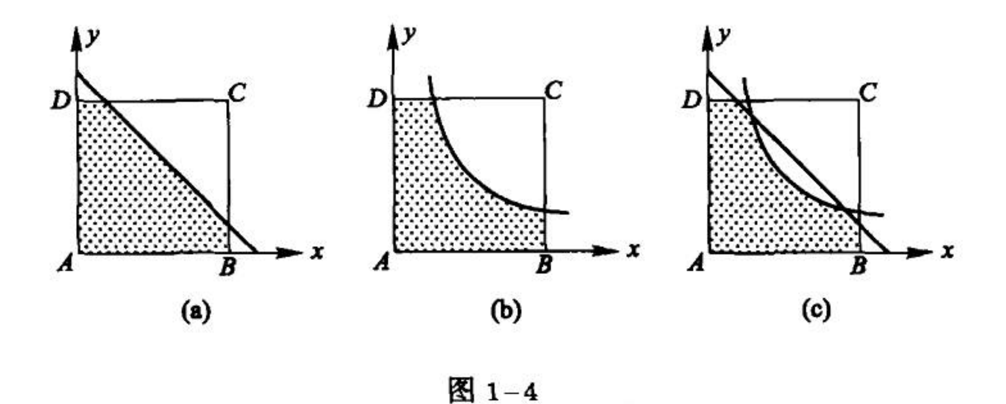


图 1-4

(1) 直线 $x+y=1.2$ 与 BC 交点坐标为 $(1,0.2)$, 与 DC 交点坐标为 $(0.2,1)$, 所以由几何概率可得

$$P=\left(1-\frac{1}{2}\times 0.8\times 0.8\right)\Big/1=0.68$$

(2) 双曲线 $xy=0.25$ 与 BC 交点坐标为 $(1,0.25)$, 与 DC 交

点坐标为 $(0.25, 1)$, 所以由几何概率可得

$$P = 0.25 \times 1 + \int_{0.25}^{1} \frac{0.25}{x}\,\mathrm{d}x = 0.596\,6$$

(3) 直线 $x + y = 1.2$ 与双曲线 $xy = 0.25$ 的交点坐标为 $(0.932, 0.268)$ 和 $(0.268, 0.932)$, 所以由几何概率可得

$$\begin{aligned} P &= 0.2 \times 1 + \int_{0.2}^{0.268} (1.2 - x)\,\mathrm{d}x \\ &\quad + \int_{0.268}^{0.932} \frac{0.25}{x}\,\mathrm{d}x + \int_{0.932}^{1} (1.2 - x)\,\mathrm{d}x \\ &= 0.593 \end{aligned}$$

【评注】几何概率典型题目之一.

25. 从 $(0, 1)$ 中随机地取二数 b 及 c, 试求方程 $x^2 + bx + c = 0$ 有实根的概率.

提示: 有实根的充要条件是判别式 $\Delta = b^2 - 4ac \geqslant 0$, 此处 $a = 1$ (图 1-5).

答 $\frac{1}{12}$.

【评注】几何概率典型题目之二.

26. 在一张打上方格的纸上投一枚直径为 1 的硬币, 方格要多小才能使硬币与线不相交的概率小于 1%.

解 设方格边长为 a, 当硬币圆心落于图 1-6 中阴影部分才与边界不相交, 由几何概率得不相交的概率为 $P = \frac{(a-1)^2}{a^2}$.

因为当 $a \leqslant 1$ 时, 硬币必定与线相交, 故只需考虑 $a > 1$, 由 $\frac{(a-1)^2}{a^2} < 0.01$ 得 $a < \frac{10}{9}$.

【评注】取材于一种美国游戏.

27. 某码头只能容纳一只船, 现预知某日将独立来到两只船, 且在 24 小时内各时刻来到的可能性都相等, 如果它们需要停靠的时间分别为 3 小时及 4 小时, 试求有一船要在江中等待的概率.

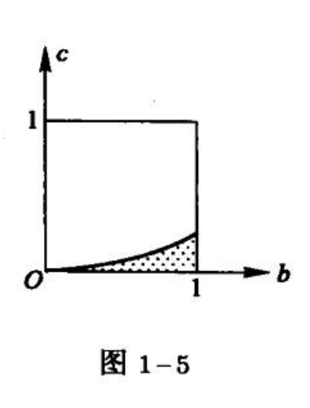


图 1-5

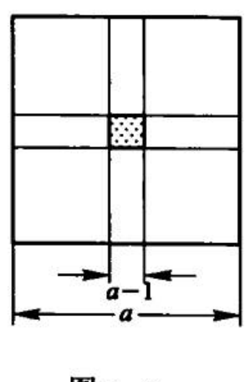


图 1-6

提示: 仿会面问题 (图 1-7).

答 $\dfrac{311}{1\ 152}$.

【评注】会面问题推广之一.

28. 两人约定于 7 点到 8 点在某地会面,试求一人要等另一人半小时以上的概率.

提示: 半小时内会不到面 (图 1-8).

答 $\dfrac{1}{4}$.

【评注】会面问题推广之二.

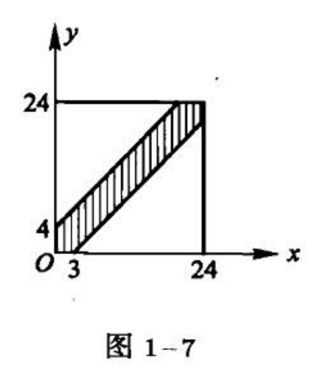


图 1-7

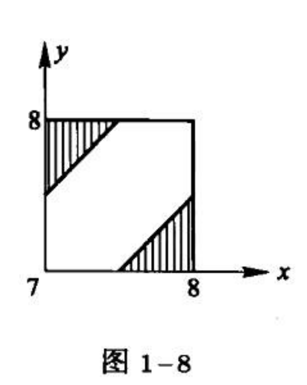


图 1-8

29. 在一线段 AB 中随机地取两个点把线段截为三段, 求这三段可以构成一个三角形的概率 (三线段能构成三角形的充要条件是任意二边之和大于第三边).

解 设在线段 AB 中随机取两个点, 从左到右依次记为 X_1 和 X_2, 令 $AB=a$, $AX_1=x_1$, $AX_2=x_2$, 则有 $0\leqslant x_1\leqslant x_2\leqslant a$. (x_1,x_2) 与三角形 EFG 内的点是一一对应的 (图 1–9).

三线段构成三角形的充要条件是

$$\begin{cases} x_1+(x_2-x_1)>a-x_2, \\ x_1+(a-x_2)>(x_2-x_1), \\ (x_2-x_1)+(a-x_2)>x_1 \end{cases}$$

这决定三角形区域 PQR, 故所求概率为 $\dfrac{1}{4}$.

【评注】又一类几何概率典型问题.

30. 在线段 $[0,1]$ 上任意投三个点, 问由 0 至三点的三线段能构成三角形与不能构成三角形这两个事件中哪一个事件的概率大.

解 设 0 到三点的三线段长分别为 x, y, z, 记相应的右端点的坐标为 x, y, z, 显然有 $0\leqslant x,y,z\leqslant 1$. 这三条线段构成三角形的充要条件是

$$x+y>z,\quad y+z>x,\quad z+x>y$$

在线段 $[0,1]$ 上任意投三点 x, y, z 与立方体 $0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$, $0\leqslant z\leqslant 1$ 中的点 (x,y,z) 一一对应, 可见所求 "构成三角形" 的概率问题可归结为在边长为 1 的立方体 T 中均匀地掷点, 而点落在区域

$$x+y>z,\quad y+z>x,\quad z+x>y$$

中, 即相当于落在四面体

$$O\text{-}ABC \text{ 和 } D\text{-}ABC$$

中 (图 1–10), 易见相应的体积是

$$V(T)=1,\quad V(O\text{-}ABC)+V(D\text{-}ABC)=\frac{1}{2}$$

由此得, "能构成三角形" 与 "不能构成三角形" 两个事件的概率一样大.

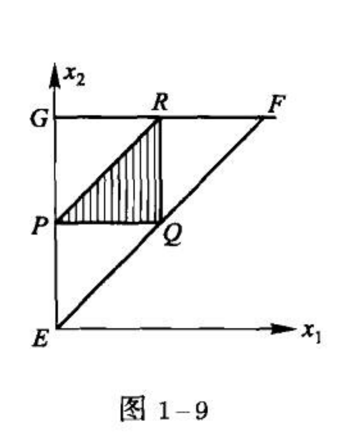


图 1-9

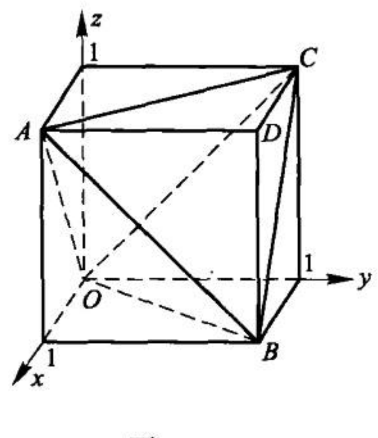


图 1-10

【评注】几何概率的又一典型题.

总的说, 几何概率题型有限. 本书正文与习题已显示其中大部分.

*31. 从一只装有 100 只灯泡的箱子中任抽 5 只灯泡, 发现有 2 只是次品, 你对此批灯泡的次品数作何估计? (这种抽查当然用不放回方式. 比较用最大似然估计法所得结果与用频率估计概率法的结果是否相同.)

解 设 100 只灯泡中有次品灯泡 n 只.

频率估计概率法:

$$\widehat{n} = \frac{2}{5} \times 100 = 40$$

最大似然估计法:

$$p_n = \frac{\dbinom{100-n}{3}\dbinom{n}{2}}{\dbinom{100}{5}}$$

$$\frac{p_n}{p_{n-1}} = \frac{n(98-n)}{(101-n)(n-2)}$$

可以算得当 $n = \dfrac{202}{5}$ 时, $\dfrac{p_n}{p_{n-1}} = 1$, $n < \dfrac{202}{5}$ 时, 则 $p_n > p_{n-1}$,

$n > \dfrac{202}{5}$ 时, 则有 $p_n < p_{n-1}$, 因为 n 取整数, 所以当 $\widehat{n} = 40$ 时, 有 p_{40} 最大.

因此据最大似然原理当 100 只灯泡中含 40 只次品时, 任抽 5 只发现 2 只次品的可能性最大.

本题中两类方法估计的结果是一致的.

【评注】数理统计中参数估计的两种主要方法的预示.

*32. 利用概率论的想法证明下列恒等式:

$$1 + \frac{A-a}{A-1} + \frac{(A-a)(A-a-1)}{(A-1)(A-2)} + \cdots + \frac{(A-a)\cdots 2\cdot 1}{(A-1)\cdots(a+1)a} = \frac{A}{a}$$

其中 A, a 都是正整数, 且 $A > a$.

证 设袋中有 A 只球, 其中有 a 只是白球, 其余为黑球. 现不放回地从袋中逐个取球, 则第 k 次才首次取得白球的概率为

$$p_1 = \frac{a}{A}, \quad p_k = \frac{a(A-a)(A-a-1)\cdots(A-a-k+2)}{A(A-1)(A-2)\cdots(A-k+1)},$$

$$k = 2, \cdots, A-a+1$$

因为袋中只有 a 只白球, 其余为黑球, 所以第 1 次或第 2 次 …… 或至多到第 $A-a+1$ 次必取到白球, 因此必有

$$\begin{aligned} 1 &= p_1 + p_2 + \cdots + p_{A-a+1} \\ &= \frac{a}{A} + \frac{a(A-a)}{A(A-1)} + \cdots + \frac{a(A-a)\cdots 2\cdot 1}{A(A-1)\cdots(a+1)a} \end{aligned}$$

等式两边同乘以 $\dfrac{A}{a}$ 得

$$1 + \frac{A-a}{A-1} + \frac{(A-a)(A-a-1)}{(A-1)(A-2)} + \cdots + \frac{(A-a)\cdots 2\cdot 1}{(A-1)\cdots(a+1)a} = \frac{A}{a}$$

【评注】这种类型的问题甚多, 证明思路基本一致, 但模型变化多端, 仅举一例作示范.

33. 设 $A_1, A_2, \cdots, A_n$ 是随机事件, 试用归纳法证明下列公式:

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i<j \leqslant n} P(A_i A_j)$$

$$+ \sum_{1 \leqslant i<j<k \leqslant n} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n)$$

提示: 从 $P(A \cup B) = P(A) + P(B) - P(AB)$ 出发.

【评注】这联系于数学中的一个基本公式, 它在不同分支以不同形式出现, 以不同的发现者命名. 证法甚多, 本题指定用归纳法证明, 组合学中用容斥法证明, 参看教学札记之四 "一般加法公式及其推广",《概率论基础》习题四 1 题提供另一种证明思路.

34. 某班有 N 个士兵, 每人各有一支枪, 这些枪外形完全一样, 在一次夜间紧急集合中, 若每人随机地取走一支枪, 问至少有一个人拿到自己的枪的概率.

解 设 A_i={第 i 个士兵拿到自己的枪}, $i = 1, 2, \cdots, N$,

$$P(A_i) = \frac{(N-1)!}{N!} = \frac{1}{N}$$

$$P(A_i A_j) = \frac{(N-2)!}{N!} = \frac{1}{N(N-1)}, \quad i \neq j$$

$$\cdots\cdots\cdots\cdots$$

$$P(A_1 A_2 \cdots A_N) = \frac{1}{N!}$$

由 33 题的公式得 "至少有一个士兵拿到自己的枪" 的概率为

$$P(A_1 \cup A_2 \cup \cdots \cup A_N)$$

$$= \sum_{i=1}^{N} P(A_i) - \sum_{1 \leqslant i<j \leqslant N} P(A_i A_j) + \cdots + (-1)^{N-1} P(A_1 A_2 \cdots A_N)$$

$$= \binom{N}{1}\frac{1}{N} - \binom{N}{2}\frac{1}{N(N-1)} + \cdots + (-1)^{N-1}\binom{N}{N}\frac{1}{N!}$$

$$= 1 - \frac{1}{2!} + \cdots + (-1)^{N-1}\frac{1}{N!}$$

$$= \sum_{k=1}^{N}\frac{(-1)^{k-1}}{k!}$$

【评注】这是蒙特莫特 (Montmort) 问题 (匹配问题) 的另一种表述, 常见的变形还有: 两副扑克牌的配对, 聚会中糊涂侍者错发衣帽, 交换礼品时拿回自己送出的东西等.

*35. 在上题中求恰好有 $k\,(0 \leqslant k \leqslant N)$ 个人拿到自己的枪的概率.

解 在 N个士兵中任意选出 k 个士兵来有 $\binom{N}{k}$ 种选法.

一组指定的 k 个士兵都拿到自己的枪的概率为

$$\frac{1}{N(N-1)\cdots(N-k+1)}$$

由 34 题 $N-k$ 个士兵都没有拿到自己的枪的概率为

$$1 - \sum_{l=1}^{N-k}\frac{(-1)^{l-1}}{l\,!} = \sum_{l=0}^{N-k}\frac{(-1)^{l}}{l\,!}$$

于是, 恰好有 k 个士兵拿到自己枪的概率为

$$P_{[k]} = \binom{N}{k}\frac{1}{N(N-1)\cdots(N-k+1)}\sum_{l=0}^{N-k}\frac{(-1)^{l}}{l\,!} = \frac{1}{k!}\sum_{l=0}^{N-k}\frac{(-1)^{l}}{l\,!}$$

【评注】上题的一般化. 算出一个完整的分布.

36. 从一副扑克牌中有放回地一张张抽取, 求在抽取第 6 张时得到全部 4 种花色的概率.

解 以 A_1, A_2, A_3, A_4 分别记抽得的 6 张牌中有黑桃、红心、

方块、草花的事件，则得到全部 4种花色的概率为

$$
\begin{aligned}
P(A_1A_2A_3A_4) &= 1 - P(\overline{A}_1 \cup \overline{A}_2 \cup \overline{A}_3 \cup \overline{A}_4) \\
&= 1 - \sum_{i=1}^{4} P(\overline{A}_i) + \sum_{1 \leqslant i < j \leqslant 4} P(\overline{A}_i \overline{A}_j) \\
&\quad - \sum_{1 \leqslant i < j < k \leqslant 4} P(\overline{A}_i \overline{A}_j \overline{A}_k) + P(\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4) \\
&= 1 - 4\left(\frac{3}{4}\right)^6 + 6\left(\frac{2}{4}\right)^6 - 4\left(\frac{1}{4}\right)^6 \\
&= 0.380\ 86
\end{aligned}
$$

【评注】利用对偶原理化交为并，巧用一般加法公式. 本题是一类古典概型典型问题中较容易的一个.

*37. 考试时共有 N 张考签，n 个学生参加考试 $(n \geqslant N)$，被抽过的考签立刻放回，求在考试结束之后，至少有一张考签没有被抽到的概率.

解 不妨将考签编号为 $1, 2, \cdots, N$，记事件 A_i={第 i 号考签未被抽到}，则

$$P(A_i) = \frac{(N-1)^n}{N^n}, \quad i = 1, 2, \cdots, N$$

$$P(A_iA_j) = \frac{(N-2)^n}{N^n}, \quad i \neq j, \quad i, j = 1, 2, \cdots, N$$

$$\cdots\cdots\cdots\cdots$$

$$P(A_1A_2\cdots A_N) = \frac{(N-N)^n}{N^n} = 0$$

易知，诸 A_i 相容，则可用本章习题 33 的公式，得

P{至少有一张考签未被抽到}

$$
\begin{aligned}
&= P(A_1 \cup A_2 \cup \cdots \cup A_N) \\
&= \sum_{i=1}^{N} P(A_i) - \sum_{1 \leqslant i < j \leqslant N} P(A_iA_j) + \cdots + (-1)^{N-1} P(A_1A_2\cdots A_N)
\end{aligned}
$$

$$=\binom{N}{1}\frac{(N-1)^n}{N^n}-\binom{N}{2}\frac{(N-2)^n}{N^n}+\cdots+(-1)^{N-2}\binom{N}{N-1}\frac{1}{N^n}+0$$

$$=\sum_{i=1}^{N-1}(-1)^{i-1}\binom{N}{i}\frac{(N-i)^n}{N^n}$$

【评注】适用一般加法公式的题型之一. 若改求每张考签都被抽到的概率, 你将如何处理?

*38. (赠券收集) 食品厂把印有水浒 108 将之一的画卡作为赠券装入某种儿童食品袋中, 每袋一卡, 试求购买 n 袋这种食品而能收齐全套画卡的概率.

解 记 B={购买的 n 袋食品袋中含有全套 108 将的画卡}, A_i={购买 n 袋食品已收集到第 i 张画卡}, $i=1,2,\cdots,108$, 则

$$B=\bigcap_{i=1}^{108}A_i,\qquad \overline{B}=\bigcup_{i=1}^{108}\overline{A}_i$$

并且

$$P(\overline{A}_i)=\frac{(108-1)^n}{108^n}=\left(1-\frac{1}{108}\right)^n$$

$$P(\overline{A}_i\,\overline{A}_j)=\frac{(108-2)^n}{108^n}=\left(1-\frac{2}{108}\right)^n,\quad i\neq j$$

$$\cdots\cdots\cdots\cdots$$

$$P(\overline{A}_1\,\overline{A}_2\cdots\overline{A}_{108})=\frac{(108-108)^n}{108^n}=0$$

因而

$$\begin{aligned}P(\overline{B})&=P(\overline{A}_1\cup\overline{A}_2\cup\cdots\cup\overline{A}_{108})\\&=\sum_{i=1}^{108}P(A_i)-\sum_{1\leqslant i<j\leqslant 108}P(\overline{A}_i\,\overline{A}_j)+\cdots-P(\overline{A}_1\,\overline{A}_2\cdots\overline{A}_{108})\\&=\binom{108}{1}\left(1-\frac{1}{108}\right)^n-\binom{108}{2}\left(1-\frac{2}{108}\right)^n+\cdots+\binom{108}{107}\left(1-\frac{107}{108}\right)^n\\&=\sum_{k=1}^{108}(-1)^{k-1}\binom{108}{k}\left(1-\frac{k}{108}\right)^n\end{aligned}$$

于是

$$P(B)=1-P(\overline{B})$$
$$=1-\sum_{k=1}^{108}(-1)^{k-1}\binom{108}{k}\left(1-\frac{k}{108}\right)^n$$
$$=\sum_{k=0}^{108}(-1)^k\binom{108}{k}\left(1-\frac{k}{108}\right)^n$$

【评注】经典问题, 变形甚多. 思考本题与前两题的联系.

*39. 用概率论想法求 N 阶行列式的展开式中包含主对角线元素的项数.

解 设 N 阶行列式中元素为 a_{ij}, 行列式展开式的每一项为不同行不同列元素的乘积, 共有 $N!$ 项.

设 A_k 表示事件 {随机选一项其中含主对角线元素 a_{kk}}, $k=1,2,\cdots,N$, 则

$$P(A_k)=\frac{(N-1)!}{N!}=\frac{1}{N}$$

$$P(A_iA_j)=\frac{(N-2)!}{N!}=\frac{1}{N(N-1)},\quad (i\neq j)$$

$$\cdots\cdots\cdots\cdots$$

$$P(A_1A_2\cdots A_N)=\frac{1}{N!}$$

任取一项含主对角线元素的概率为

$$P(A_1\cup A_2\cup\cdots\cup A_N)$$
$$=\sum_{i=1}^{N}P(A_i)-\sum_{1\leqslant i<j\leqslant N}P(A_iA_j)+\cdots+(-1)^{N-1}P(A_1A_2\cdots A_N)$$
$$=\binom{N}{1}\frac{1}{N}-\binom{N}{2}\frac{1}{N(N-1)}+\cdots+(-1)^{N-1}\binom{N}{N}\frac{1}{N!}$$

$$
=1-\frac{1}{2!}+\cdots+(-1)^{N-1}\frac{1}{N!}
$$

$$
=\sum_{k=1}^{N}\frac{(-1)^{k-1}}{k!}
$$

由此得包含主对角线元素的项数为

$$
N!\sum_{k=1}^{N}\frac{(-1)^{k-1}}{k!}
$$

【评注】思路来自主对角线元素的两个足标是相互“匹配”的.

以上数题紧密联系一般加法公式, 关于这个主题, 可参考教学札记之四“一般加法公式及其推广”.

40. 有 w 只白球与 b 只黑球任你放入两个袋子中, 让你的朋友随机抽一袋并从中摸出一只球, 你将如何做以使你的朋友摸得黑球的概率最大.

答 一个袋放一只黑球, 另一个袋放入 w 只白球和 $b-1$ 只黑球.

【评注】直观可得答案, 证明较难. (提示: 把 $k \leqslant \frac{w+b}{2}$ 只球放入一袋, 其中有 x 只黑球, 写出摸得黑球的概率, 再讨论 x 取何值时使上述概率最大.)

*41. 甲, 乙, 丙三人按下面规则进行比赛, 第一局由甲, 乙参加而丙轮空, 由第一局的优胜者与丙进行第二局比赛, 而失败者则轮空, 比赛用这种方式一直进行到其中一个人连胜两局为止, 连胜两局者成为整场比赛的优胜者, 若甲, 乙, 丙胜每局的概率各为 $\frac{1}{2}$, 问甲, 乙, 丙成为整场比赛优胜者的概率各是多少?

解 若以 A、B、C 分别表示甲、乙、丙成为整场的优胜者, 以 a、b、c 分别表示在一局中甲胜、乙胜和丙胜, 则整场比赛的可能结果可表示为:

$$
aa,\quad acc,\quad acbb,\quad acbaa,\quad acbacc,\quad acbacbb,\cdots
$$

$$bb,\quad bcc,\quad bcaa,\quad bcabb,\quad bcabcc,\quad bcabcaa,\cdots$$

于是有

$$\begin{aligned}P(C) &= \big[P(acc)+P(bcc)\big]+\big[P(acbacc)+P(bcabcc)\big]+\cdots \\ &= 2\times\frac{1}{2^3}+2\times\frac{1}{2^6}+2\times\frac{1}{2^9}+\cdots \\ &= \frac{2}{7}\end{aligned}$$

由于甲、乙两人所处的地位是对称的, 所以有 $P(A)=P(B)$, 故得

$$P(A)=P(B)=\frac{1}{2}\left(1-\frac{2}{7}\right)=\frac{5}{14}$$

【评注】事件分析与概率计算训练题之一.

*42. 父, 母, 子三人举行比赛, 每局总有一人胜一人负 (没有和局), 每局的优胜者就与未参加此局的人再进行比赛, 如果某人首先胜了两局, 则他就是整个比赛的优胜者, 由父决定第一局由哪两人参加, 其中儿子实力最强, 所以父为了使自己得胜的概率达到最大, 就决定第一局由他与妻子先比赛, 试证父的决策为最优策略 (任何一对选手中一人胜对方的概率在整个比赛中是不变的) .

证 单局比赛有六种情况, 记父胜子的事件为 F_S, 相应概率为 p_{F_S}. 对应地, 子胜父记为 S_F, 概率记为 p_{S_F}; 父胜母记为 F_M, 概率记为 p_{F_M}; 母胜父记为 M_F, 概率记为 p_{M_F}; 母胜子记为 M_S, 概率记为 p_{M_S}; 子胜母记为 S_M, 概率记为 p_{S_M}. 易见 $p_{M_S}+p_{S_M}=1$, $p_{F_S}<p_{F_M}$. 若按父的决策首局由父与母比赛, 则整个比赛的可能结果为

$$\begin{aligned}&F_MF_S,\quad F_MS_FS_M,\quad F_MS_FM_SM_F,\quad F_MS_FM_SF_M\\ &M_FM_S,\quad M_FS_MS_F,\quad M_FS_MF_SF_M,\quad M_FS_MF_SM_F\end{aligned}$$

故有

$$\begin{aligned}P_1 &= P\{\text{第一局父对母而父为整个比赛优胜者}\} \\ &= P(F_MF_S)+P(F_MS_FM_SF_M)+P(M_FS_MF_SF_M) \\ &= p_{F_M}p_{F_S}+p_{F_M}p_{S_F}p_{M_S}p_{F_M}+p_{M_F}p_{S_M}p_{F_S}p_{F_M}\end{aligned}$$

类似地有

$$P_2 = P\{第一局父对子而父为整个比赛优胜者\}$$

$$= P(F_S F_M) + P(F_S M_F S_M F_S) + P(S_F M_S F_M F_S)$$

$$= p_{F_S} p_{F_M} + p_{F_S} p_{M_F} p_{S_M} p_{F_S} + p_{S_F} p_{M_S} p_{F_M} p_{F_S}$$

$$P_3 = P\{第一局子对母而父为整个比赛优胜者\}$$

$$= P(S_M F_S F_M) + P(M_S F_M F_S)$$

$$= p_{S_M} p_{F_S} p_{F_M} + p_{M_S} p_{F_M} p_{F_S}$$

$$= p_{F_S} p_{F_M} (p_{S_M} + p_{M_S})$$

$$= p_{F_S} p_{F_M}$$

经比较易知 $P_3 < P_2$. 又由于 $p_{F_S} < p_{F_M}$, 所以,

$$p_{F_S} p_{M_F} p_{S_M} p_{F_S} < p_{M_F} p_{S_M} p_{F_S} p_{F_M}$$

$$p_{S_F} p_{M_S} p_{F_M} p_{F_S} < p_{F_M} p_{S_F} p_{M_S} p_{F_M}$$

因此有 $P_2 < P_1$, 从而 $P_1 > P_2 > P_3$, 这说明父的决策是最优策略.

【评注】中学数学奥林匹克赛题. 复杂事件分析与概率计算.

43. 给定 $p = P(A)$, $q = P(B)$, $r = P(A \cup B)$, 求 $P(A\overline{B})$ 及 $P(\overline{A}\,\overline{B})$.

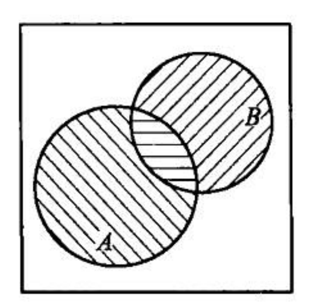


图 1-11

提示: 从文图出发.

答 $P(A\overline{B}) = r - q$,

$P(\overline{A}\,\overline{B}) = 1 - r$.

【评注】本题是举例性的, 事实上给定三个概率 p, q, r 即能算出有关 A, B 的全部 16 个事件 (见本章习题 46 题) 的概率. 思考为什么? 进一步可考虑推广到三个事件 A, B, C 的场合.

44. 设 p_1, p_2, p_{12} 是给定的实数, 试证存在两个事件 A_1 及 A_2

使得 $P(A_1)=p_1$, $P(A_2)=p_2$, $P(A_1A_2)=p_{12}$ 的充要条件是下列四个不等式同时成立:

$$p_{12}\geqslant 0,\quad p_1-p_{12}\geqslant 0,\quad p_2-p_{12}\geqslant 0,\quad 1-p_1-p_2+p_{12}\geqslant 0$$

证 必要性. 设存在 A_1 及 A_2 使得 $P(A_1)=p_1$, $P(A_2)=p_2$, $P(A_1A_2)=p_{12}$, 则

$$0\leqslant p_1\leqslant 1,\quad 0\leqslant p_2\leqslant 1,\quad 0\leqslant p_{12}\leqslant 1$$

$$0\leqslant P(A_1-A_1A_2)=P(A_1)-P(A_1A_2)=p_1-p_{12}$$

$$0\leqslant P(A_2-A_2A_1)=P(A_2)-P(A_1A_2)=p_2-p_{12}$$

$$1\geqslant P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1A_2)=p_1+p_2-p_{12}$$

所以有

$$p_{12}\geqslant 0,\quad p_1-p_{12}\geqslant 0,\quad p_2-p_{12}\geqslant 0,\quad 1-p_1-p_2+p_{12}\geqslant 0$$

充分性. 由第一个不等式知

$$p_{12}\geqslant 0$$

再由另三个不等式得

$$p_{12}\geqslant p_1+p_2-1\geqslant p_{12}+p_{12}-1$$

故又知

$$p_{12}\leqslant 1$$

而

$$0\leqslant p_1-p_{12}\leqslant 1-p_2\leqslant 1-p_{12}\leqslant 1$$

类似得

$$0\leqslant p_2-p_{12}\leqslant 1$$

再用第四个不等式知道存在三个互不相容的事件 B, C, D 使

$$P(B)=p_1-p_{12},\quad P(C)=p_{12},\quad P(D)=p_2-p_{12}$$

于是可令

$$A_1=B+C,\quad A_2=C+D$$

即有

$$P(A_1) = p_1 - p_{12} + p_{12} = p_1$$

$$P(A_2) = p_2 - p_{12} + p_{12} = p_2$$

$$P(A_1A_2) = P(C) = p_{12}$$

【评注】事件域中诸事件的概率的给定要受各种约束.

45. 证明: $|P(AB) - P(A)P(B)| \leqslant \dfrac{1}{4}$, 并讨论等号成立的条件.

证

$$\begin{aligned}&P(A)P(B) - P(AB)\\&= P(A)\left[P(AB) + P(\overline{A}B)\right] - P(AB)\\&= P(A)P(\overline{A}B) - P(AB)\left[1 - P(A)\right]\\&\leqslant P(A)P(\overline{A}B) \leqslant P(A)\left[1 - P(A)\right] \leqslant \frac{1}{4}\end{aligned}$$

另一方面 (不妨设 $P(A) \geqslant P(B)$),

$$\begin{aligned}&P(AB) - P(A)P(B)\\&\leqslant P(B) - P(B)P(B) = P(B)\left[1 - P(B)\right] \leqslant \frac{1}{4}\end{aligned}$$

因此

$$\left|P(AB) - P(A)P(B)\right| \leqslant \frac{1}{4}$$

从上述证明可看到, 要使 $P(A)\left[1 - P(A)\right] \leqslant \dfrac{1}{4}$ 中的等号成立当且仅当 $P(A) = \dfrac{1}{2}$, 由对称性知 $P(B) = \dfrac{1}{2}$.

由此可得出当且仅当 $P(A) = P(B) = \dfrac{1}{2}$, 且 $P(AB) = 0$ 即 $B = \overline{A}$ 时, 或者 $P(AB) = \dfrac{1}{2}$ 即 $B = A$ 时所求证的不等式的等号成立.

【评注】第四章 §2 将再回到这个不等式, 并提供新见解.

46. 求包含事件 A, B 的最小 σ 域.

解 数字 0, 1 可以表示两种不同的状态.

若以 1 表示样本点 ω 属于某事件, 而 0 则表示不属于, 则对两个事件 A, B 会有 $2^2=4$ 种情况, 它们分别代表四个事件, 这四个事件互不相容, 就如同下表所示:

A	B		
1	1	表示	AB
1	0	表示	$A\overline{B}$
0	1	表示	$\overline{A}B$
0	0	表示	$\overline{A}\,\overline{B}$

现在考虑这四个互不相容事件 $AB, A\overline{B}, \overline{A}B, \overline{A}\,\overline{B}$, 这四个互不相容事件构成了全空间 Ω 的一个分割. 为了找出最小 σ 域, 即其全体子集, 考虑事件 "并" 的运算, 现在以 1 表示该事件参与 "并" 的运算, 而 0 则表示不参与, 这样就有 $2^4=16$ 种情况, 列于下表:

AB	$A\overline{B}$	$\overline{A}B$	$\overline{A}\,\overline{B}$		
0	0	0	0	表示	$\varnothing$
1	0	0	0	表示	AB
0	1	0	0	表示	$A\overline{B}$
0	0	1	0	表示	$\overline{A}B$
0	0	0	1	表示	$\overline{A}\,\overline{B}$
1	1	0	0	表示	A
1	0	1	0	表示	B
1	0	0	1	表示	$AB\cup\overline{A}\,\overline{B}$
0	1	1	0	表示	$A\overline{B}\cup\overline{A}B$
0	1	0	1	表示	$\overline{B}$
0	0	1	1	表示	$\overline{A}$
1	1	1	0	表示	$A\cup B$
1	1	0	1	表示	$A\cup\overline{B}$
1	0	1	1	表示	$\overline{A}\cup B$
0	1	1	1	表示	$\overline{A}\cup\overline{B}$
1	1	1	1	表示	Ω

因此, 该[illegible] σ 域为 $\{\Omega, \varnothing, A, B, \overline{A}, \overline{B}, AB, A\overline{B}, \overline{A}B, \overline{A}\,\overline{B}, AB\cup\overline{A}\,\overline{B}, A\overline{B}\cup\overline{A}B, A\cup B, \overline{A}\cup\overline{B}, \overline{A}\cup B, A\cup\overline{B}\}$.

【评注】由事件 A, B 把样本空间分割成 4 个不相容事件 $A\overline{B}$, $\overline{A}B$, AB, $\overline{A}\,\overline{B}$ 之和, 由它们生成的 $2^4=16$ 个事件构成上述 σ 域.

前面三题讨论了在这个事件域给定概率时会遇到的某些问题.

请读者推广到三个事件 A, B, C 的场合.

47. 证明: (1) Ω 的一切子集组成的集类是一个 σ 域; (2) σ 域之交仍为 σ 域.

证 (1) 记 $\mathscr{F}=\{\Omega$ 的一切子集$\}$.

(i) Ω 是 Ω 的子集, 所以 $\Omega\in\mathscr{F}$;

(ii) 若 $A\in\mathscr{F}$, 即 A 是 Ω 的子集, 则 $\overline{A}=\Omega-A$ 也是 Ω 的子集, 所以 $\overline{A}\in\mathscr{F}$;

(iii) 若 $A_i\in\mathscr{F}$, $i=1,2,\cdots$, 则有 $A_i\subset\Omega$. 对任一 $\omega\in\bigcup\limits_{i=1}^{\infty}A_i$, 必有某一 i_0, 使 $\omega\in A_{i_0}\subset\Omega$, 从而有 $\bigcup\limits_{i=1}^{\infty}A_i\subset\Omega$, 即 $\bigcup\limits_{i=1}^{\infty}A_i$ 也是 Ω 的一个子集, 故 $\bigcup\limits_{i=1}^{\infty}A_i\in\mathscr{F}$.

所以 $\mathscr{F}$ 是 σ 域.

(2) 设 $\mathscr{F}_t\,(t\in T)$ 是 σ 域, 记 $\mathscr{F}=\bigcap\limits_{t\in T}\mathscr{F}_t$.

(i) $\Omega\in\mathscr{F}_t$, $\forall t\in T$, 所以 $\Omega\in\bigcap\limits_{t\in T}\mathscr{F}_t$, 即 $\Omega\in\mathscr{F}$;

(ii) 若 $A\in\mathscr{F}$, 则 $A\in\mathscr{F}_t$, $\forall t\in T$. 由 $\mathscr{F}_t$ 是 σ 域得 $\overline{A}\in\mathscr{F}_t$, $\forall t\in T$, 所以 $\overline{A}\in\bigcap\limits_{t\in T}\mathscr{F}_t$, 从而有 $\overline{A}\in\mathscr{F}$;

(iii) 若 $A_i\in\mathscr{F}$, $i=1,2,\cdots$, 则 $A_i\in\mathscr{F}_t$, $\forall t\in T$. 由于 $\mathscr{F}_t$ 是 σ 域, 所以 $\bigcup\limits_{i=1}^{\infty}A_i\in\mathscr{F}_t$, $\forall t\in T$, 即 $\bigcup\limits_{i=1}^{\infty}A_i\in\bigcap\limits_{t\in T}\mathscr{F}_t=\mathscr{F}$.

所以 $\mathscr{F}$ 是 σ 域.

【评注】这是《概率论基础》中用到的两个重要结论,特留给学生作为练习,有助于记牢 σ 域的三个要求.

48. 证明: 包含一切形为 $(-\infty,\ x)$ 的区间的最小 σ 域是一维博雷尔 σ 域.

证 一维博雷尔 σ 域 $\mathscr{B}=\sigma\{[a,b)\}$ 是由左闭右开区间类产生的 σ 域.

设 $\widetilde{\mathscr{B}}=\sigma\{(-\infty,x)\}$ 是由形如 $(-\infty,x)$ 区间类产生的 σ 域.

因为 $[a,b)=(-\infty,b)-(-\infty,a)$, $\widetilde{\mathscr{B}}$ 是 σ 域, 所以 $[a,b)\in\widetilde{\mathscr{B}}$, 因此有 $\mathscr{B}\subset\widetilde{\mathscr{B}}$.

又由于 $(-\infty,x)=\bigcup\limits_{n=1}^{\infty}[x-n,x-n+1)$, $\mathscr{B}$ 是 σ 域, 所以 $(-\infty,x)\in\mathscr{B}$, 因此有 $\widetilde{\mathscr{B}}\subset\mathscr{B}$.

于是有 $\widetilde{\mathscr{B}}=\mathscr{B}$.

【评注】这也是《概率论基础》中反复用到的一个论断, 通过完成本题, 学生将对博雷尔 σ 域 $\mathscr{B}$ 有进一步了解.

49. (1) 设 Q 是定义在 σ 域上的非负广义实值函数 (即可以取有限或无限值的函数), 如果它具有可列可加性, 并且 $Q(\varnothing)=0$, 则称 Q 为**测度**, 试说明测度概念是算术中计数概念及几何中长度、面积、体积等概念的推广; (2) 用测度概念解释古典概型、几何概率及概率论公理化结构中关于概率的定义.

答 (1) 由 $Q(\varnothing)=0$ 及测度的可列可加性易得测度具有有限可加性.

算术中的计数. 以 $n(E)$表示集合 E 中包含的元素的个数.

(i) $n(E)$ 非负;

(ii) 若 $E_i,\ i=1,2,\cdots,l$ 中任意两个 E_i 与 $E_j\ (i\neq j)$ 之间都没有相同的元素, 则 $n(\sum\limits_{i=1}^{l}E_i)=\sum\limits_{i=1}^{l}n(E_i)$, 即集合的和包含的元素的个数等于各个集合包含元素个数的和, 也就是说 $n(E)$ 具有有

限可加性;

(iii) 若 $E=\varnothing$, 它不包含任何元素, 则有 $n(\varnothing)=0$.

几何度量中的长度. 以 $m(E)$ 表示区间 E 的长度.

(i) $m(E)$ 非负;

(ii) 对区间 E_i, $i=1,2,\cdots$, 若任意两个 E_i 与 E_j $(i\neq j)$ 都不相交, 则 $m(\sum\limits_{i=1}^{\infty}E_i)=\sum\limits_{i=1}^{\infty}m(E_i)$, $m(E)$ 具有可列可加性;

(iii) 空集 $\varnothing$ 的长度 $m(\varnothing)=0$.

当区间改成平面或空间区域, 长度改成面积或体积时, 以上结论仍成立.

非负性, 可列可加性, 空集对应的值为零, 这些是算术中计数概念和几何学中度量衡概念所共有的性质, 因此可以看到测度概念是上述这些概念的推广.

(2) 古典概型中概率的定义: 设 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$, $\forall A\subset\Omega$, 定义其概率为 $P(A)=\dfrac{A\text{ 包含的样本点数}}{\Omega\text{ 中样本点总数}}$. P 具有非负性, 有限可加性, $P(\varnothing)=0$. 此处的概率 P 相当于算术中的比例, 只是 P 还有规范性, 即 $P(\Omega)=1$. 古典概型中的概率实质上是定义在 $\mathscr{F}=\{\Omega\text{ 的一切子集}\}$ 上的具有规范性的测度.

几何概率的定义: 设 Ω 与 $\forall G\subset\Omega$ 都是 (n 维欧几里得空间的) 可测集. 对 G 定义其概率为 $P(G)=\dfrac{m(G)}{m(\Omega)}$, 其中 $m(G)$ 表示区域 G 的测度, 可见 P 具有非负性, 可列可加性, 此外 $P(\varnothing)=0$, $P(\Omega)=1$, 所以 P 是定义在 Ω 中可测集上的具有规范性的测度.

在概率论公理化结构中, 定义在事件域 $\mathscr{F}$ 上的集合函数 P, 若满足

(i) 非负性: $P(A)\geqslant 0,\quad \forall A\in\mathscr{F}$;

(ii) 规范性: $P(\Omega)=1$;

(iii) 可列可加性: 若 $A_i\in\mathscr{F}$, $i=1,2,\cdots$, $A_iA_j=\varnothing$, $i\neq j$, 则

$$P\Big(\sum_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i),$$

那么称 P 为 $\mathscr{F}$ 上的概率, 由这三条性质可推得 $P(\varnothing) = 0$, 因此可知定义在 $\mathscr{F}$ 上的概率 P 实质上就是定义在 $\mathscr{F}$ 上的规范性测度.

【评注】简直可以看作本章的小结!

*50. 试证: 概率定义1.5.2中的三个要求可用下列两个要求代替:

(i) $P(A) \geqslant 0$, 对一切 $A \in \mathscr{F}$;

(ii) 若 $A_i \in \mathscr{F}$, $i = 1, 2, \cdots$, 两两互不相容, 且 $\sum_{i=1}^{\infty} A_i = \Omega$, 则

$$\sum_{i=1}^{\infty} P(A_i) = 1.$$

证 概率定义的三个条件为

① 非负性: $P(A) \geqslant 0, \quad \forall A \in \mathscr{F}$;

② 规范性: $P(\Omega) = 1$;

③ 可列可加性: 若 $A_i \in \mathscr{F}$, $i = 1, 2, \cdots$, $A_i A_j = \varnothing$, $i \neq j$, 则

$$P\Big(\sum_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i)$$

显然 ①与 (i) 是等价的.

若 $A_i \in \mathscr{F}$, $i = 1, 2, \cdots$, 两两互不相容, 且 $\sum_{i=1}^{\infty} A_i = \Omega$, 则由

②③可推得 $1 = P(\Omega) = P\Big(\sum_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i)$, 即 (ii) 成立.

反之, 当 (ii) 成立, 则由 $\Omega = \Omega + \varnothing + \varnothing + \cdots$ 可得 $P(\varnothing) = 0$, $P(\Omega) = 1$, 因此 ②成立, 再由 $\Omega = B + \overline{B} + \varnothing + \varnothing + \cdots$ 可得 $P(B) + P(\overline{B}) = 1$. 若 $\forall A_i \in \mathscr{F}$, $i = 1, 2, \cdots$, 且两两互不相容, 则因 $\sum_{i=1}^{\infty} A_i + \overline{\sum_{i=1}^{\infty} A_i} = \Omega$, 由 (ii) 可得 $\sum_{i=1}^{\infty} P(A_i) + P\Big(\overline{\sum_{i=1}^{\infty} A_i}\Big) = 1$, 加

上前证 $P\Big(\sum\limits_{i=1}^{\infty}A_i\Big)+P\Big(\overline{\sum\limits_{i=1}^{\infty}A_i}\Big)=1$, 所以有 $P\Big(\sum\limits_{i=1}^{\infty}A_i\Big)=\sum\limits_{i=1}^{\infty}P(A_i)$, 即 ③可列可加性成立.

【评注】介绍公理备择形式, 训练学生论证能力.

习题总评

第一章习题的训练重点是事件分析与概率计算, 主要结合古典概型与几何概率进行. 古典概型约 20 题, 前三分之二用直接法, 包括众多不同类型; 后三分之一用公式法, 大体是一些名题. 几何概率共计 7 题, 也包括了常见的类型. 总的说这两类习题都有一定难度, 不是懂得定义就能做出来的, 而是充满技巧性, 对学生而言, 是有趣而艰难的训练.

其他题目则紧扣教材, 完成它们有助于掌握本章的有关知识.

章后小议

本章最后给出了概率论这一学科最基本的两个概念 —— 事件与概率的严格定义.

事件的定义建立在集合论的基础上, 而概率则经过细致的考察后定义成规范化测度.

三个具体概率模型被仔细研究. 1° 同样条件下的大量重复试验 (尚欠严格); 2° 古典概型; 3° 几何概率.

有两种办法讲授概率的定义:

第一种, 立刻给出概率的公理化定义, 然后作为特例讲古典概型、几何概率以及其他模型. 优点是直入主题, 节省时间.

另一种则是《概率论基础》采用的方式: 从古典概型、几何概率等具体模型入手, 最后引导到近代的公理化体系. 这种做法可以让学生在简单场合感受概率直观和得到更多解题技巧训练并增

添课程的趣味性, 也使本学科对青年学生更有吸引力.

从古典概型、几何概率、统计定义出发讲概率论, 最初见于格涅坚科《概率论教程》.《概率论基础》继承这个体系并有所发挥.

鉴于概率的统计定义有诸多缺陷,《概率论基础》有意回避统计定义一说, 但强调频率稳定性, 并以三大实验性例子为证, 说明概率的客观性, 同时把频率与概率的关系贯穿全书.

对古典概型及几何概率,《概率论基础》从正文到习题都有较齐全的配置.

为定义概率论的基本概念, σ 域, 特别是欧几里得空间上的博雷尔 σ 域是不可或缺的概念, 书中只对其作适可而止的处理.

教学札记之一

浅谈集合的大小

现代概率论建立在集合论的基础上. 样本空间 Ω 是样本点的全体构成的集合; 事件域 $\mathscr{F}$ 定义为样本空间的某些子集构成的集类; 概率则定义为 $\mathscr{F}$ 上的规范化测度.

概率问题的复杂程度及其所使用的数学工具的高低都跟所要处理的样本点的集合的大小有关系, 因此样本空间实际上是按集合的大小归类的——有限, 可列, $\mathbf{R}^1$, $\mathbf{R}^2$, $\mathbf{R}^n$ ······ 等等.

概率论处理的样本空间, 多数是无穷集合, 因此有必要对集合论的相关内容作些介绍.

集合论于 19 世纪 70 年代由德国数学家康托尔 (G.Cantor, 1845—1918) 创立, 现已成为数学各学科的公共基础.

集合论研究 “无穷大”, 也使人类对实数的理解大大深入.

无穷大的概念困惑数学家与哲学家达 2000 年之久, 到康托尔才获得突破, 得出许多令人意外的结论.

从亚里士多德开始, 只承认 $\{1,2,3,\cdots\}$ 的 “潜无穷”, 不敢承认 $[0,1]$ 中的点的 “实无穷”, 直至 19 世纪上半叶, 甚至大数学家

柯西、高斯等人也不敢对无穷大作深入探讨.

一些现象令人费解. 例如自然数可与其平方数一一对应, 但后者明显少于前者; 又如从圆心向两同心圆画出半径可建立两圆周上点的一一对应, 但大圆的长度显然大于小圆; 最简单的, $y = 2x$ 把 $[0, 1]$ 的点一一对应于 $[0, 2]$ 的点.

在有限的场合比较大小, 显然局部小于整体, 但在无穷的场合, 并不如此.

康托尔用一一对应来比较集合的大小.

在有限场合, 集合的大小能用某一个自然数来表示, 与常识完全一致.

若 Ω 有 n 个元素, 则它有 $\binom{n}{k}$ 个包含 k 个元素的子集, 还包含一个空集, 一个全集, 也都算子集, 因此其子集全体构成的集类共有 $\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}=2^n$ 个元素, 也还是有限集.

在无穷场合, 则发现有各种无穷大, 康托尔用势 (power) 来表示.

1. 记自然数的势为 $\aleph_0$ (读阿列夫零), 此为第一级无穷大, 称为可列或可数.

2. 偶数与自然数同势.

3. 整数也与自然数同势.

4. 可列个可列仍为可列.

图 1–12 中可列个可列集中的元素依对角线顺序重新排列即可.

5. 有理数是可列的.

有理数表示成分数, 以图 1–13 方式排列 (可能有重复), 按对角线顺序重新排列即可.

6. 实数全体是不可列的.

$[0,1]$ 中的每个实数可用无穷小数 $0.a_1a_2a_3a_4\cdots$ 来表示 (为使表示法唯一, 规定 0.3 表示为 $0.2999\cdots$, 余类推). 为证明 $[0,1]$ 中的实数是不可列的, 用反证法. 设 $[0,1]$ 中实数全体可排列成

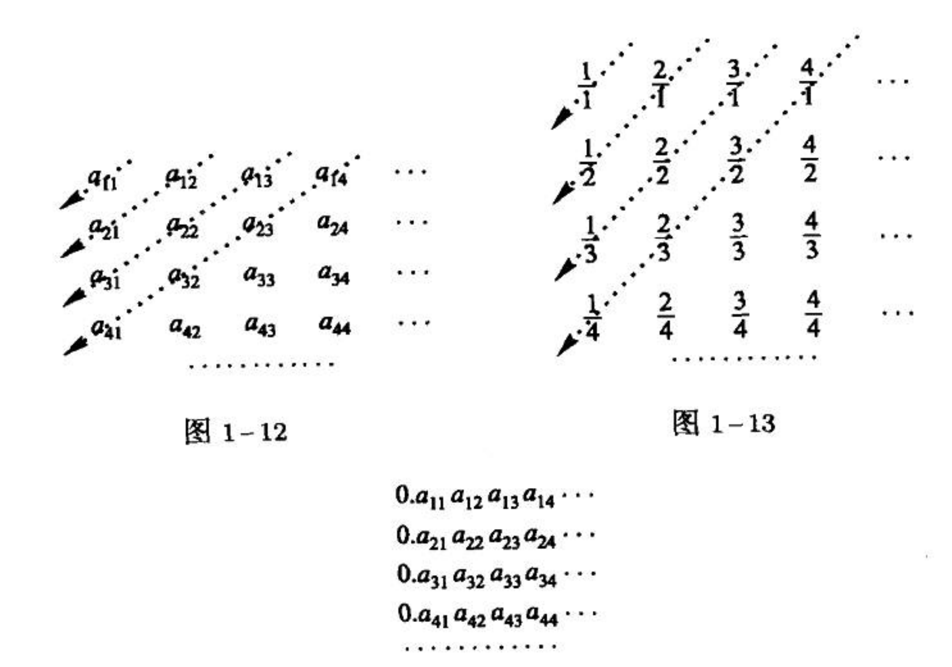


图 1－12　　　　图 1－13

$$
\begin{array}{l}
0.a_{11}a_{12}a_{13}a_{14}\cdots \\
0.a_{21}a_{22}a_{23}a_{24}\cdots \\
0.a_{31}a_{32}a_{33}a_{34}\cdots \\
0.a_{41}a_{42}a_{43}a_{44}\cdots \\
\cdots\cdots\cdots
\end{array}
$$

取 $b=0.b_1b_2b_3b_4\cdots$, 其中 $b_i \neq a_{ii}$, 显然 b 为 $[0,1]$ 中实数, 但与上列各数都不同, 得出矛盾. 故证得 $[0,1]$ 中实数全体是不可列的.

$[0,1]$ 中的点可用 $y=\tan\left(x-\dfrac{1}{2}\right)\pi$ 同 $(-\infty,\infty)$ 建立一一对应, 故知 $\mathbf{R}^1$ 是不可列的. $\mathbf{R}^1$ 的势记为 C, 称为**连续统**, 显然它是比 $\aleph_0$ 更高级的无穷大.

7. 无理数全体不可列.

8. 二维欧几里得空间 $\mathbf{R}^2$ 的势也是 C.

只需注意到

$$
x=0.a_1a_2a_3a_4\cdots \left\langle \begin{array}{l} x_1=0.a_1a_3\cdots \\ x_2=0.a_2a_4\cdots \end{array}\right.
$$

建立 $[0,1]$ 与 $[0,1]\times[0,1]$ 的一一对应.

9. n 维欧几里得空间 $\mathbf{R}^n$ 的势也是 C.

10. $2^{\aleph_0}=C$.

$[0,1]$ 中的数可用二进制表示为 $0.b_1b_2b_3b_4\cdots$, 其中 $b_i=\begin{cases}0,\\1,\end{cases}$

其势为 C; 另一方面, 这个表达式又可以理解为一个含有可列个元素的集合当它的第 i 个元素选入子集时 b_i 用 1 表示, 不选入子集用 0 表示, 0, 1 的所有选法对应其子集全体, 故它构成的集类之势为 $2^{\aleph_0}$, 这二者一一对应.

11. 集合 A 的一切子集之势高于 A 的势. 因此势也是无穷的.

概率论只讨论到样本空间的一切子集所构成的集类这一级, 不算太难.

基础概率论中用到的样本空间 Ω:

(1) 有限样本空间, Ω 的势为 n, 其一切子集之势为 2^n.

(2) 可列样本空间, Ω 的势为 $\aleph_0$, 其一切子集之势为 $2^{\aleph_0}=C$.

(3) 欧几里得空间 $\mathbf{R}^1$, $\mathbf{R}^2$, $\mathbf{R}^n$, Ω 的势为 C, 讨论的子集类限于博雷尔 σ 域.

在有限样本空间, 概率论所用数学工具只是初等数学, 在可列样本空间还用级数, 但在欧几里得空间则广泛使用微积分的各种知识.

教学札记之二

排列组合辑要

一、排列、组合公式

1. 乘法原理: 完成一件工作需经两个过程, 进行第一过程有 n_1 种方法, 第二过程有 n_2 种方法, 则完成该工作共有 $n_1 \times n_2$ 种方法. 显然可把上述结果推广到多个过程的场合.

下列公式的推导只基于这个原理.

2. 排列: 从 n 个不同元素中任意选出 r 个元素排成一列. 既考虑元素, 也考虑顺序.

重复排列数 (r 个元素不相同、有部分相同或全部相同):

选排列数 (r 个元素均不相同):

$$A_n^r = n(n-1)\cdots(n-r+1), \quad 1 \leqslant r \leqslant n$$

选排列的特例: $r=n$ 时称为全排列,

$$P_n = n!, \quad \text{定义 } 0! = 1$$

显然有 $A_n^r = \dfrac{n!}{(n-r)!}$. 当 $r>n$ 时, 定义 $A_n^r = 0$.

3. 组合: 从 n 个不同元素中任意选出 r 个元素. 只考虑元素, 不考虑顺序.

组合数 (r 个元素均不相同):

$$\binom{n}{r} = \begin{cases} \dfrac{n(n-1)\cdots(n-r+1)}{r!}, & 1 \leqslant r \leqslant n \\ 0 &, \quad r > n \end{cases}$$

定义 $\binom{n}{0} = 1$.

重复组合数 (r 个元素中可以有相同的):

$$H_n^r = \binom{n+r-1}{r}$$

4. 小结: 排列、组合公式归纳如下

	重复	不重复
排列	n^r	A_n^r
组合	$\binom{n+r-1}{r}$	$\binom{n}{r}$

排列与组合的分类不能绝对化, 例如两类元素的全排列等价于组合.

二、两类典型问题

1. 抽样 (摸球): 从 n 张卡片中抽取 r 张.

抽样有两种方式: 放回, 不放回.

对结果有两种记录: 讲序 (有序), 不讲序 (无序).

从 n 张卡片中抽取 r 张可能结果总数列于下表:

	放回	不放回
有序	n^r	A_n^r
无序	$\binom{n+r-1}{r}$	$\binom{n}{r}$

2. 分配 (占位): 把 r 只球投入 n 个房间.

每个房间容球的规则: 不限制, 有限制 (只限一球).

球的个性: 可分辨 (编号), 不可分辨.

r 只球投入 n 个房间可能结果总数列于下表:

	不限制	有限制
可分辨	n^r	A_n^r
不可分辨	$\binom{n+r-1}{r}$	$\binom{n}{r}$

把 r 只球投入 n 个房间还有另一种表述.

若以向量 $(n_1, n_2, \cdots, n_r)$ 记第 i 只球投入的房间号码为 n_i, $i = 1, 2, \cdots, r$, 并称为房号向量. 再以向量 $(r_1, r_2, \cdots, r_n)$ 记第 j 号房间落入的球数为 r_j, $j = 1, 2, \cdots, n$, 并称为球数向量, 则可能的向量总数列于下表:

	不限制	有限制
房号向量	n^r	A_n^r
球数向量	$\binom{n+r-1}{r}$	$\binom{n}{r}$

三、重复排列

$$\underbrace{n \cdot n \cdot \cdots \cdot n}_{r} = n^r$$

1. 中国人很早研究有重复排列与二进制:《易传 · 系辞上传》: "易有太极, 是生两仪, 两仪生四象, 四象生八卦. "传说周文王演八

卦, 成六十四卦.

2. n 位二进制数即 $(x_1, x_2, \cdots, x_n)$, 其中 $x_i = \begin{cases} 0, \\ 1, \end{cases}$ 计有 2^n 个.

3. 含有 n 个元素的集合的子集总数为 2^n 个.

4. 彩票摇奖器当众摇出一个 6 位数的中奖号码 (可从 000 000 到 999 999), 共有 10^6 种可能结果.

5. 掷一颗骰子 4 次, 共有 6^4 种可能结果, 掷两颗骰子 24 次共有 36^{24} 种可能结果.

6. 从装有编号为 1 到 n 的球的袋子中有放回地摸出 r 只球, 依次记下其号码, 可能总数为 n^r.

7. n 个人的生日 (以每年 365 天计) 共有 365^n 种.

8. 公共汽车上的 r 位乘客在以后的 n 个站下车的方式计有 n^r 种.

9. (麦克斯韦 – 玻尔兹曼统计) n 个可分辨粒子在相空间的 N 个相格中的可能分配有 N^n 种.

四、全排列

$$P_n = n(n-1)\cdots 3\cdot 2\cdot 1$$

1. n 卷文集在书架上的排列方式有 $n!$ 种.

2. 把 n 个客人分到 n 个房间共有 $n!$ 种安排法.

3. 把编号为 1 到 N 的球从袋中一只接一只摸出依序排成一列, 共有 $N!$ 种不同结果.

4. 把 n 封信分别装入 n 只信封共有 $n!$ 种结果.

五、选排列

$$A_n^r = (n)_r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}, \quad 1 \leqslant r \leqslant n$$

1. 从 n 种牙膏中选 r 种依序上货架, 共有 A_n^r 种不同排列.

2. 把 r 个客人分到 n $(n \geqslant r)$ 个房间, 共有 A_n^r 种安排法.

3. 把 n 只可分辨的球投入 N 个房间, 若每个房间只容一球, 且 $n \leqslant N$, 则有 A_N^n 种分配法.

六、组合

$$\binom{n}{r} = \mathrm{C}_n^r = \frac{A_n^r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!\,(n-r)!}$$

主要性质:

杨辉恒等式: $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$.

图 1–14 摘自杨辉《详解九章算法》(1261), 该书注明 "贾宪用此术". 贾宪是 11 世纪前半叶北宋数学家. 帕斯卡于 1654 年完成论文《论算术三角形》, 故西方称为帕斯卡三角或算术三角形.

定义 $\binom{n}{0} = 1$, 则

$$\binom{n}{r} = \binom{n}{n-r}, \quad 0 \leqslant r \leqslant n$$

```
          1
        1   1
      1   2   1
    1   3   3   1
  1   4   6   4   1
1   5  10  10   5   1
......................
```

图 1–14

二项式: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$

令 $a = b = 1$, 得 $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n$, 再次表明含有 n 个元素的集合的子集总数为 2^n.

常用公式:

$$\binom{n+1}{r} = \binom{n}{r-1} + \binom{n-1}{r-1} + \binom{n-2}{r-1} + \cdots + \binom{r-1}{r-1}$$

1. (抽样) 从 n 个不同元素中取 r 个不计顺序, 其总数为 $\binom{n}{r}$.

2. (分割) 把 n 个物体分为两类, 第一类 r 个, 第二类 $n-r$ 个,

分法共有 $\binom{n}{r}$ 种.

3. (两类元素的全排列) n 个元素中有 r 个是一类, $n-r$ 是另一类, 则不同的全排列数为 $\dfrac{n!}{r!(n-r)!}=\binom{n}{r}$ 种.

4. (二项系数) 二项展开式 $(1+x)^n=\sum\limits_{k=0}^{n}\binom{n}{k}x^k$ 中 x^r 项的系数为 $\binom{n}{r}$.

5. (占位) r 个黑球在排成一行的 n 个位置上有 $\binom{n}{r}$ 种放法.

6. 连投 n 次篮球有 r 次命中共有 $\binom{n}{r}$ 种方式实现.

7. (费米－狄拉克统计) r 个不可分辨的粒子在 n 个量子态 (每个量子态至多有一个粒子占据) 中的分配数为 $\binom{n}{r}$.

七、多项系数

$$\binom{n}{r_1,r_2,\cdots,r_k}=\frac{n!}{r_1!r_2!\cdots r_k!},\quad r_1+r_2+\cdots+r_k=n$$

当 $k=2$ 时化为二项系数.

1. (抽样) 从 n 个不同的元素中取 r_1 个构成第 1 组, r_2 个构成第 2 组, ……, 剩下的 r_k 个构成第 k 组, 计有 $\dfrac{n!}{r_1!r_2!\cdots r_k!}$ 种.

2. (分割) 把 n 个物体分为 k 类, 第一类 r_1 个, 第二类 r_2 个, ……, 第 k 类 r_k 个, 这里 $r_1+r_2+\cdots+r_k=n$, 则分法共有 $\dfrac{n!}{r_1!r_2!\cdots r_k!}$ 种.

3. (k 类元素的全排列) n 个元素中有 r_1 个是第一类, r_2 个是第二类, ……, r_k 个是第 k 类, $r_1+r_2+\cdots+r_k=n$, 则可分辨的全排列数为 $\dfrac{n!}{r_1!r_2!\cdots r_k!}$.

4. (多项系数) 多项展开式

$$(x_1+x_2+\cdots+x_k)^n=\sum_{\substack{r_1+r_2+\cdots+r_k=n\\0\leqslant r_i\leqslant n}}\frac{n!}{r_1!r_2!\cdots r_k!}x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$$

中 $x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$ 项的系数为 $\dfrac{n!}{r_1!r_2!\cdots r_k!}$.

八、重复组合

记从 n 个不同的元素里, 每次取出 r 个元素的有重复的组合的总数为 H_n^r, 则

$$H_n^r=\binom{n+r-1}{r}$$

1. 多项展开式:

$$(a_1+a_2+\cdots+a_n)^r=\sum_{\substack{r_1+r_2+\cdots+r_n=r\\0\leqslant r_i\leqslant r}}\frac{r!}{r_1!r_2!\cdots r_n!}a_1^{r_1}a_2^{r_2}\cdots a_n^{r_n}$$

共有 $\dbinom{n+r-1}{r}$ 个同类项.

特别地, $(a+b)^r=\sum\limits_{0\leqslant k\leqslant r}\dbinom{r}{k}a^kb^{r-k}$ 共有 $r+1$ 个 (同类) 项.

证明 (归纳法)

$n=1$, $H_1^r=\dbinom{r}{r}=1$.

$n=2$, $H_2^r=\dbinom{r+1}{r}=r+1$, 即上述二项式.

若 $H_{n-1}^r=\dbinom{n+r-2}{r}$, 则 H_n^r 种选法中, 从原 $n-1$ 个元素中选出的共有 $H_{n-1}^r=\dbinom{n+r-2}{r}$ 种, 余为加入新元素, 新元素可以出 $1,2,\cdots,r$ 次. 当新元素出 i 次时, 另外 $r-i$ 次从原 $n-1$

个元素选出, 计有 H_{n-1}^{r-i} 种, $i=1,2,\cdots,r$, 故

$$\begin{aligned}H_n^r &= H_{n-1}^r + H_{n-1}^{r-1} + H_{n-1}^{r-2} + \cdots + H_{n-1}^1 + 1\\ &= \binom{n+r-2}{r}+\binom{n+r-3}{r-1}+\binom{n+r-4}{r-2}+\cdots+\binom{n-1}{1}+1\\ &= \binom{n+r-2}{n-2}+\binom{n+r-3}{n-2}+\binom{n+r-4}{n-2}+\cdots+\binom{n-1}{n-2}+1\\ &= \binom{n+r-1}{n-1}=\binom{n+r-1}{r}\end{aligned}$$

从而证得结论.

2. 满足不定方程 $x_1+x_2+\cdots+x_n=r, 0\leqslant x_i\leqslant r$, $i=1,2,\cdots,n$ 的非负整数解 $(x_1,x_2,\cdots,x_n)$ 的组数为 H_n^r.

证明 (0-1 序列)

当 x_i 取非负整数 k 就用连排的 k 个 1 表示, 加号用 0 表示, 则方程化为 r 个 1 与 $n-1$ 个 0 的排列, 每种排列对应方程的一组解, 由两类元素的全排列数知计有 $\binom{n+r-1}{r}$ 组, 从而证得结论.

3. 从 $1,2,3,\cdots,n$ 中有放回地选 r 个数, 计有 H_n^r 种选法.

证明 (重复抽样)

把选出的数依大小排成 $a_1\leqslant a_2\leqslant a_3\leqslant\cdots\leqslant a_r$, 当然其中可能有些数相等. 作如下变换: 从第二个数起分别加 $1,2,3,\cdots,r-1$, 即得

$$\begin{array}{cccccc} & a_1, & a_2, & a_3, & \cdots, & a_r\\ + & & 1, & 2, & \cdots, & r-1\\ \hline & b_1, & b_2, & b_3, & \cdots, & b_r\end{array}$$

显然 $b_1,b_2,b_3,\cdots,b_r$ 不等.

从 $1,2,3,\cdots,n$ 中有放回地选一组数 $a_1,a_2,a_3,\cdots,a_r$ 与从 $1,2,3,\cdots,n,n+1,\cdots,n+r-1$ 中不放回地选一组数 $b_1,b_2,b_3,\cdots,b_r$ 是一一对应的, 后者共有 $\binom{n+r-1}{r}$ 种选法, 从而证得结论.

4. (玻色－爱因斯坦统计) r 个不可分辨的粒子在 n 个无限制的相格中的分配总数为 H_n^r.

证明 (投球入格)

若以 0 表示球, 用 $||$ 表示房间, 则 r 个球在排成一列的 n 个房间中的分配相当于 r 个球与 $n-1$ 根棒的排列 (因为可以省去两端的棒), 选定球的位置或棒的位置都可确定一种排列法, 计有

$$\binom{n+r-1}{r}=\binom{n+r-1}{n-1}$$

种选法, 从而证得结论.

九、k 类元素的抽样数

1. 袋中有 M 只黑球, $N-M$ 只白球, 从中摸出 n 只, 则有 k 只黑球、$n-k$ 只白球的构成方法计有 $\binom{M}{k}\binom{N-M}{n-k}$ 种.

2. 袋中装有 i 号球 N_i 只, $i=1,2,\cdots,k$, $N_1+N_2+\cdots+N_k=N$, 从中摸出 n 只, 则有 i 号球 n_i 只, $i=1,2,\cdots,k$, $n_1+n_2+\cdots+n_k=n$ 的构成方法有 $\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_k}{n_k}$ 种.

3. 若 $r_1+r_2+\cdots+r_k=n$, 则

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\binom{n-r_1-r_2}{r_3}\cdots\binom{n-r_1-r_2-\cdots-r_{k-1}}{r_k}$$
$$=\frac{n!}{r_1!r_2!\cdots r_k!}$$

教学札记之三

对称性与概率计算

归根到底, 概率论肇始于如下认识: 对称事件其发生的可能性相等. 因此发掘对称性是计算概率的一种特殊技巧, 下面是一些例子.

一、n 只球在 n 个位子的全排列

(1) 1 号球出现在 1 号位的概率为 $\frac{1}{n}$, 由对称性, 1 号球出现在第 k 号位的概率也为 $\frac{1}{n}$, 更进一步, 第 i 号球出现在第 k 号位的概率也为 $\frac{1}{n}$. 这是不放回开锁问题的答案.

(2) k 号位由预先指定的某号球占据的概率为 $\frac{1}{n}$.

(3) 袋中有 a 只黑球 b 只白球, 从中一只只摸出球, 则第 k 次摸得黑球的概率为 $\frac{a}{a+b}$, 与顺序无关, 这是抽签问题的答案.

(4) 在 n 封信与 n 只信封的匹配中, 1 号信装入 1 号信封的概率为 $\frac{1}{n}$, 由对称性, i 号信装入 i 号信封的概率也为 $\frac{1}{n}$;

1, 2 号信与信封符合的概率为 $\frac{1}{n}\cdot\frac{1}{n-1}$, 由对称性, 第 i, j 号信与信封符合的概率也为 $\frac{1}{n(n-1)}$;

1, 2, 3 号信与信封符合的概率为 $\frac{1}{n}\cdot\frac{1}{n-1}\cdot\frac{1}{n-2}$, 故第 i, j, k 号信与信封符合的概率也为 $\frac{1}{n(n-1)(n-2)}$.

……

以上结果用于匹配问题的解题中.

(5) 袋中有 a 只黑球和 b 只白球, 从袋中每次摸出一只球, 直到袋中剩下的球全是同色球为止, 求袋中剩下的球全是黑球的概率.

解 这是题集中常见的一道 "烦题", 事实上其答案一目了然.

记 A={剩下的球全是黑球}, B={最后一只球是黑球}, 不难验证 $A=B$, 由对称性得 $P(B)=\frac{a}{a+b}$, 故 $P(A)=\frac{a}{a+b}$.

二、抽样调查

(1) 从 N 张卡片中任取一张, 1 号或 k 号卡片被抽到的概率均为 $\frac{1}{N}$.

(2) 从 N 张卡片中任取两张, 1, 2 号卡片被抽到的概率为 $\frac{1}{N} \cdot \frac{1}{N-1} \times 2$, 由对称性, 预先指定的 k 号与 j 号卡片被抽到的概率亦为 $\frac{1}{\binom{N}{2}}$.

(3) 从 N 张卡片中任取 l 张, 预先指定的 $i_1, i_2, \cdots, i_l$ 号卡片被抽到的概率为 $\frac{1}{\binom{N}{l}} = \frac{l!}{N(N-1)\cdots(N-l+1)}$, 这也表明, 在抽样中, l 张卡片一张张被抽出与一次性抽出 l 张卡片是等价的.

上述结论在抽样调查的计算中经常隐性使用.

(4) 袋中有 a 只黑球 b 只白球, 从中摸出 n 只, 有 k 只黑球的概率为 $\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}$, 即超几何分布.

(5) 甲袋中有 a 只黑球 b 只白球, 乙袋和丙袋都是空袋, 先从甲袋中随机摸 $n\,(1 \leqslant n \leqslant a+b)$ 只球放入乙袋, 再从乙袋中随机摸 $m\,(1 \leqslant m \leqslant n)$ 只球放入丙袋, 最后从丙袋中任取一球, 求该球为黑球的概率.

这也是题集中的一道. 没有任何一只球比别的球更有机会成为最后从丙袋中取出的球, 由对称性可知答案为 $\frac{a}{a+b}$.

三、掷硬币

(1) 甲有 $n+1$ 枚硬币, 乙有 n 枚, 求甲掷出的正面数超过乙掷出的正面数的概率.

解 设乙掷出 k 个正面, 甲掷出 $k+l$ 个正面, $0\leqslant k\leqslant n$, $l\geqslant 1$, 则

$$
\begin{aligned}
P\{\text{甲正} > \text{乙正}\} &= \frac{1}{2^{2n+1}}\sum_{l=1}^{n+1}\sum_{k=0}^{n}\binom{n+1}{k+l}\binom{n}{k} \\
&= \frac{1}{2^{2n+1}}\sum_{l=1}^{n+1}\sum_{k=0}^{n}\binom{n+1}{n+1-k-l}\binom{n}{k} \\
&= \frac{1}{2^{2n+1}}\sum_{l=1}^{n+1}\binom{2n+1}{n+1-l} \\
&= \frac{1}{2^{2n+1}}\sum_{i=0}^{n}\binom{2n+1}{i} \\
&= \frac{1}{2^{2n+1}}\cdot\frac{1}{2}\sum_{i=0}^{2n+1}\binom{2n+1}{i} \\
&= \frac{1}{2^{2n+1}}\cdot\frac{1}{2}\cdot 2^{2n+1} = \frac{1}{2}
\end{aligned}
$$

如果熟悉组合系数的计算, 这个通用解法也不能算太难, 但利用对称性则能很快得到答案.

记 $A=\{\text{甲正} > \text{乙正}\}$, $B=\{\text{甲反} > \text{乙反}\}$, 显然 A 与 B 对称, 故 $P(A) = P(B)$.

在甲只比乙多一枚硬币的条件下 (只有这种特殊情况下才有此特殊解法!), $\overline{B} = A$, 因此又有 $P(A) + P(B) = 1$, 即得 $P(A) = \frac{1}{2}$.

1965 年我为复旦大学数学系 67 届数学专业两个小班上概率论课, 同学们提出约十种不同的解法得出本题的答案 $\frac{1}{2}$, 多数正确, 少数错误. 其中的一种错误做法是: 前 n 个对等, 因此甲能否超出决定于最后一枚硬币, 它出正面的概率为 $\frac{1}{2}$.

(2) 掷硬币 $2n+1$ 次, 求出现的正面数多于反面数的概率.

解 问题不同, 但有相似的解法.

记 A={正面数 > 反面数}, B={反面数 > 正面数}, 显然 $\overline{B}=A$, 另一方面由对称性知 $P(A)=P(B)$, 故 $P(A)=\dfrac{1}{2}$.

(3) 甲有 $n+2$ 枚硬币, 乙有 n 枚硬币, 投掷后比较, 求 A={甲正 > 乙正} 的概率.

解 若记 B={甲反 > 乙反}, 由对称性仍有

$$P(A)=P(B)$$

但这里 $AB\neq\varnothing$, 不过由于 $\overline{A}\,\overline{B}=\varnothing$, 故知 $P(A\cup B)=P(\Omega)=1$, 又

$$\begin{aligned}AB&=\{\text{甲正}>\text{乙正},\ \text{甲反}>\text{乙反}\}\\&=\{\text{甲正}-\text{乙正}=1,\ \text{甲反}-\text{乙反}=1\}\\&=\{\text{甲正}-\text{乙正}=1\}\end{aligned}$$

故

$$\begin{aligned}P(AB)&=\sum_{k=0}^{n}P\{\text{甲正}=k+1,\ \text{乙正}=k\}\\&=\sum_{k=0}^{n}\binom{n+2}{k+1}\binom{n}{k}\left(\frac{1}{2}\right)^{2n+2}\\&=\sum_{k=0}^{n}\binom{n+2}{n+1-k}\binom{n}{k}\frac{1}{2^{2n+2}}\\&=\binom{2n+2}{n+1}\frac{1}{2^{2n+2}}\end{aligned}$$

由

$$1=P(A)+P(B)-P(AB)$$

得到

$$P(A)=\frac{1}{2}\left[1+\frac{\dbinom{2n+2}{n+1}}{2^{2n+2}}\right]$$

四、投点排序

(1) 在线段 AB 上任投 3 点 x_1, x_2, x_3, 求 x_2 落入 x_1 与 x_3 之间的概率.

解 记 A_1={ x_1 落入 x_2 与 x_3 之间}, A_2={ x_2 落入 x_1 与 x_3 之间}, A_3={ x_3 落入 x_1 与 x_2 之间}, 显然 A_1, A_2, A_3 是对称的, 其发生的可能性相等, 又 A_1, A_2, A_3 两两不相容且 $P(A_i \cup A_2 \cup A_3) = 1$, 故 $P(A_2) = \dfrac{1}{3}$.

(2) 在线段 AB 上任投 3 点 x_1, x_2, x_3, 求事件 $\{x_1 < x_2 < x_3\}$ 发生的概率.

解 类似上题, 记

$$A_1 = \{x_1 < x_2 < x_3\}, \quad A_2 = \{x_1 < x_3 < x_2\}$$
$$A_3 = \{x_2 < x_1 < x_3\}, \quad A_4 = \{x_2 < x_3 < x_1\}$$
$$A_5 = \{x_3 < x_1 < x_2\}, \quad A_6 = \{x_3 < x_2 < x_1\}$$

它们对称且两两不相容, 又 $A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$ 发生的概率为 1, 因而

$$P(A_1) = P\{x_1 < x_2 < x_3\} = \frac{1}{6}$$

(3) 在线段 AB 上任取 n 点 x_1, x_2, $\cdots$, x_n, 则事件 $\{x_{i_1} < x_{i_2} < \cdots < x_{i_n}\}$ 发生的概率为 $\dfrac{1}{n!}$.

在数理统计中, 把 x_1, x_2, $\cdots$, x_n 推广到独立同分布场合, 这类技巧甚为有用.

五、顺序统计量

(1) 按学生名单, 以随机顺序抽出 $n+1$ 个人, 则最后一个人比其他人都高的概率, 由对称性应为 $\dfrac{1}{n+1}$.

本题也常用运动员成绩破纪录来叙述.

(2) 从学生中随意找 13 个人称体重, 依次记为 X_1, X_2, $\cdots$,

X_{13}, 求 $\max(X_1, X_2, \cdots, X_{10}) < \min(X_{11}, X_{12}, X_{13})$ 的概率.

解 依对称性看作 13 个不等数的排列, 计有 13! 种, 其中 $10! \cdot 3!$ 种满足要求, 故所求概率为

$$\frac{10!3!}{13!} = \frac{1}{286}$$

以上两题曾分别作为中国科技大学概率论与数理统计专业研究生的入学试题.

(3) 随机抽取的 n 个学生中, 第 i 号与第 j 号是最矮二位的概率为 $\dfrac{1}{\dbinom{n}{2}}$.

(4) 随机抽取的 n 个学生中, 第一个人次矮, 第 n 个人最矮的概率为 $\dfrac{1}{n(n-1)}$.

换一种说法, 对一件倒霉事 (例如金钱损失) 去问一批朋友, 直问到第 $n-1$ 个人才发现一个比自己更倒霉的, 这事发生的概率为 $\dfrac{1}{n(n-1)}$. 这被费勒称为倒霉事的持续时间, 并指出其期望值为无穷大.

(5) 甲班有 M 个男生, 乙班有 N 个男生, 把他们依身高分别排成 $X_1^{(M)} < X_2^{(M)} < \cdots < X_M^{(M)}$ 及 $Y_1^{(N)} < Y_2^{(N)} < \cdots < Y_N^{(N)}$, 求 $Y_n^{(N)} < X_{m+1}^{(M)} < Y_{n+1}^{(N)}$ $(1 \leqslant n \leqslant N-1, 0 \leqslant m \leqslant M-1)$ 的概率.

解 用对称性, $M+N$ 个身高在编号为 1, 2, $\cdots$, $M+N$ 个位子上的全排列是等概率的, 为使上式成立, $X_{m+1}^{(M)}$ 应出现在第 $n+m+1$ 个位子上, 它之前有 n 个 $Y_i^{(N)}$, m 个 $X_j^{(M)}$, 它之后有 $N-n$ 个 $Y_k^{(N)}$ 及 $M-m-1$ 个 $X_l^{(M)}$, 故所求概率为

$$p = \frac{\dbinom{n+m}{m}\dbinom{1}{1}\dbinom{M+N-n-m-1}{M-m-1}}{\dbinom{M+N}{M}}$$

教学札记之四

一般加法公式及其推广

一、一般加法公式

概率加法公式

$$P(A\cup B)=P(A)+P(B)-P(AB)$$

有非常直观的解释: 为计算事件 $A\cup B$ 的概率, 对包含在 $A\cup B$ 中的每个样本点必须计算一次, 也只需计算一次.

等式右边正是如此, 前两项把 $A\cup B$ 中的每个样本点都计算到了, 但含在 AB 中的样本点却被重复计算, 因此在第三项中扣除, 从而达到对 $A\cup B$ 中的每个样本点都精确计算一次.

在三个事件 A, B, C 的场合, 概率加法公式为

$$\begin{aligned}&P(A\cup B\cup C)\\&=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(CA)+P(ABC)\end{aligned}$$

这时, 若样本点 ω 不属于 A, B, C 中任一个, 则在右边诸项中均不出现, 未被计入左边概率: 若样本点 ω 只属于 A, B, C 中某一个, 则在右边头三项中必出现且只出现一次, 而在后边诸项中不再出现, 因此正好计算一次; 若样本点 ω 属于 A, B, C 中的某两个, 则在右边头三项中被计算了两次, 在中间三项被扣除一次, 在最后一项中不出现, 也正好计算一次; 若样本点 ω 属于 ABC, 则在右边头三项中被计算三次, 在中间三项中被扣除三次, 最后一项中又计回, 因此也同样被计算概率正好一次. 这种证明方法被称为"容斥(inclusion-exclusion)法".

对事件 $A_1, A_2, \cdots, A_n$, 记和式

$$S_1=\sum_{1\leqslant i\leqslant n}P(A_i)$$

$$S_2=\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)$$

$$S_3 = \sum_{1 \leqslant i < j < k \leqslant n} P(A_i A_j A_k)$$

$$\cdots\cdots\cdots\cdots$$

$$S_n = P(A_1 A_2 \cdots A_n)$$

一般加法公式是上述加法公式的一般化.

定理 1 (一般加法公式) 若 $A_1, A_2, \cdots, A_n$ 为 n 个事件, 则

$$P_1 = P(A_1 \cup A_2 \cup \cdots \cup A_n) = S_1 - S_2 + S_3 - \cdots + (-1)^{n-1} S_n \quad (1)$$

这是《概率论基础》的 (1.5.9) 式, 该书又把它编入习题一 33 题, 要求读者用归纳法证明. 在学过更多概率论知识之后, 在习题四 1 题中将会提示它的另一种证明思路. 下面我们将用"容斥法"来给出它的第三种证明. 如果你已读懂上述特例的证明, 那么下面的证明便不难理解.

证明 若样本点 ω 属于 $A_1, A_2, \cdots, A_n$ 中的 m $(1 \leqslant m \leqslant n)$ 个, 则在第 1 个和式中它被计算 $\binom{m}{1}$ 次, 在第 2 个和式中被计算 $\binom{m}{2}$ 次, 在第 3 个和式中被计算 $\binom{m}{3}$ 次, $\cdots\cdots$, 在第 m 个和式中被计算 $\binom{m}{m}$ 次, 在之后的和式中不再出现, 因此在 (1) 的右边一共被计入

$$\binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \cdots \pm \binom{m}{m} = 1 - (1-1)^m = 1$$

次.

二、n 个事件中正好发生 k 个的概率

定理 2 若以 B_k 记事件 $A_1, A_2, \cdots, A_n$ 中正好发生 k $(k = 0, 1, 2, \cdots, n)$ 个这一事件, 而以 C_k 记事件 $A_1, A_2, \cdots, A_n$ 中至少

发生 k 个这一事件, 则

$$\begin{aligned}P_{[k]}&=P(B_k)\\&=S_k-\binom{k+1}{k}S_{k+1}+\binom{k+2}{k}S_{k+2}-\cdots+(-1)^{n-k}\binom{n}{k}S_n\end{aligned} \tag{2}$$

这里记 $S_0=1$, 特别地

$$P_{[0]}=1-S_1+S_2-\cdots+(-1)^nS_n \tag{2'}$$

$$\begin{aligned}P_k&=P(C_k)\\&=S_k-\binom{k}{k-1}S_{k+1}+\binom{k+1}{k-1}S_{k+2}-\cdots+(-1)^{n-k}\binom{n-1}{k-1}S_n\end{aligned} \tag{3}$$

特别地

$$P_1=S_1-S_2+S_3-\cdots+(-1)^{n-1}S_n \tag{3'}$$

证明 还是用容斥法.

首先提醒, 若样本点 ω 属于 $A_1,A_2,\cdots,A_n$ 中正好 m $(1\leqslant m\leqslant n)$ 个, 则 $P\{\omega\}$ 只在 $S_1,S_2,\cdots,S_m$ 中计入, 而不会在 $S_{m+1},S_{m+2},\cdots,S_n$ 中计入. 下面进入正式证明.

当 $m=k$ 时, $P\{\omega\}$ 在且只在 (2) 的右边第一项 S_k 中被计入一次. 因此只需证明当 $m\neq k$ 时, $P\{\omega\}$ 在 (2) 的右边实际上未被计入.

事实上, 当 $m<k$ 时, $P\{\omega\}$ 不会在 $S_k,S_{k+1},\cdots,S_n$ 中计入, 因此也就不会被计入 (2) 的右边.

而当 $m>k$ 时, 由于包含 ω 的 m 个事件 $A_{i_1},A_{i_2},\cdots,A_{i_m}$ 可组成 $\binom{m}{l}$ 个 l 重交, 故 $P\{\omega\}$ 在 S_l 中被计入 $\binom{m}{l}$ 次, $l=k,k+1,\cdots,m$, 因此 $P\{\omega\}$ 在 (2) 的右边被计入的总次数为

$$N=\binom{k}{k}\binom{m}{k}-\binom{k+1}{k}\binom{m}{k+1}+\binom{k+2}{k}\binom{m}{k+2}-\cdots\pm\binom{m}{k}\binom{m}{m}$$

由于

$$\binom{k+j}{k}\binom{m}{k+j}=\binom{m}{k}\binom{m-k}{j}$$

故

$$N=\binom{m}{k}\left[\binom{m-k}{0}-\binom{m-k}{1}+\binom{m-k}{2}-\cdots\pm\binom{m-k}{m-k}\right]$$

$$=\binom{m}{k}(1-1)^{m-k}=0$$

于是证得 (2).

由定义可知

$$P_k=P_{[k]}+P_{[k+1]}+\cdots+P_{[n]} \tag{4}$$

及

$$P_{k+1}=P_k-P_{[k]} \tag{5}$$

有了 (2), 为证 (3) 可用归纳法并利用递推式 (5).

显然 $P_1=1-P_{[0]}$ 即 (3)′ 成立.

假设对 P_k 成立 (3) 式, 则利用基本组合等式

$$\binom{k+1}{k-1}-\binom{k+2}{k}=-\binom{k+1}{k}$$

由 (5)、(3)、(2) 得

$$\begin{aligned}P_{k+1}&=P_k-P_{[k]}\\&=S_{k+1}-\binom{k+1}{k}S_{k+2}+\binom{k+2}{k}S_{k+3}+\cdots+\\&\qquad(-1)^{n-k-1}\binom{n-1}{k}S_n\end{aligned}$$

这正是对应 $k+1$ 的 P_{k+1} 应成立的 (3) 式, 从而完成了归纳法所需的证明.

当然, 读者也可以把 (2) 代入 (4) 导出 (3) 来完成本定理第二部分的证明.

公式 (2) 对 $k=0,1,2,\cdots,n$ 给出一个分布, 更便于应用, 而且通过 (2)′ 也把一般加法公式 (3)′ 作为推论.

三、在匹配问题中的应用

公式 (1) 的特例最早在 1708 年由蒙特莫特对匹配问题导出，有了一般公式 (1) 之后，则匹配问题成为其应用实例之一，其表达式见于《概率论基础》第一章§5 例 6，习题一 34 题实质相同，习题一 *39 题与习题四 32 题亦有关联.

《概率论基础》没有导出公式 (2)，只是在习题一 *35 题中要求读者在匹配问题这一特例中作推导，当然这是一道有相当难度的习题.

四、在赠券收集和投球入室问题中的应用

赠券收集问题是概率论中另外一个有趣的问题，也有实用价值. 通常都假定每次购得第 $j\,(j=1,2,\cdots,n)$ 号赠券的概率均为 $\dfrac{1}{n}$，相当于把球等可能地投入 n 个房间之一，因此这两类问题，除非还有其他假定，实质上是相同的.

赠券收集问题最关心的是要集齐全套要花多少时间或金钱，这显然是概率论问题，也常用一般加法公式求解.

与上述内容关联的概率问题，《概率论基础》中均以习题的形式出现，计有习题一 36, *37, *38 题及习题四 6 题.

五、对偶公式

与一般加法公式对应的是“一般乘法公式”：

$$
\begin{aligned}
&P(A_1A_2\cdots A_n)\\
&\quad=\sum_i P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_i\cup A_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_i\cup A_j\cup A_k)\\
&\qquad-\cdots+(-1)^{n-1}P(A_1\cup A_2\cup\cdots\cup A_n) \qquad (6)
\end{aligned}
$$

这个公式可以用归纳法直接证明，也可以用事件的对偶公式利用一般加法公式来证明，还能用示性函数法证明. 不过人们通常只记得一般加法公式 (1)，它甚至被写入大部分概率论教科书，而

(6) 式却少为人知. 大概是因为凡是关于 (6) 式的计算都可以化为关于一般加法公式 (1) 的计算. 例如习题一 36 题及 *38 题就是实例.

第二章　条件概率与统计独立性

章前引言

概率论研究的主要问题之一是新信息的获得将如何影响事件发生的概率. 为此, 本章定义条件概率, 引进概率论的几个基本公式, 大大增强了处理信息的能力; 提出 (统计) 独立性概念, 形成了概率论特有的研究课题. 接着, 从伯努利试验出发, 进入学科中心, 并引入多种分布为下章作前导.

二项分布和泊松分布是概率论三大分布中的两个, 因此设专节讨论. 归结为二项分布计算的四个应用实例承前启后; 对泊松过程的讨论则留下伏笔.

课文导读

§ 2.1　条件概率, 全概率公式, 贝叶斯公式

条件概率在概率的三大背景场合 (古典概型, 几何概率, 频率) 都有直观含义, 但在一般场合只能用公式定义, 不过这些直观含义对理解条件概率还是很重要的.

条件概率用来处理信息, 扩大了概率论研究的范围和能力. 乘法公式与加法公式是早期概率论的重要成果.

全概率公式通过条件化把复杂问题分解成简单问题, 先逐个解决, 最后再解决整个问题. 这个解题思路十分普适而合理, 因而全

概率公式成为概率论中使用频率最高的公式.

贝叶斯 (1702—1761) 在他死后的 1763 年发表的一篇论文中为归纳推理提出了一整套思路, 其中包括如今以他名字命名的公式. 这套理论不管对概率论还是对统计学都产生了深远的影响, 本书只是略微涉及一点.

关于全概率公式与贝叶斯公式的应用散见在《概率论基础》的正文和众多习题中, 在教学札记之六 “全概率公式与贝叶斯公式” 中作了一个不很全面的小结, 供参考.

§ 2.2 事件独立性

本节引入 (统计) 独立性, 它是概率论中独有的概念. 有人认为, 主要是独立性让概率论从测度论的一般理论中 “独立” 出来成为一门学科.

初步的印象是独立性使许多概率计算得以简化, 它的真正重要性将逐步显现.

独立性对条件概率的影响最为明显.

概率论中有多种形式的独立性定义, 不过其实质是一样的, 最后都归根于事件的独立性, 例如试验的独立性就是一例.

事件独立性本身也有多种等价的定义方式, 搞清事件独立性的涵义是重要的.

两两独立不能推出多个也独立. 这个结论要牢记, 这只要记住伯恩斯坦的著名的反例, 也许完成习题二 18 题能提供一些更为感性的认识.

教学札记之二十三 “事件的独立性与相关性” 告诉大家, 在学习过更多知识后, 怎么从更宽广的背景上来看待事件独立性.

概率计算公式小汇

至本节, 事件概率计算公式出尽, 小结如下:

一、加法公式

$$P(A\cup B)=P(A)+P(B)-P(AB) \tag{1}$$

特例: 当 A, B 互不相容时

$$P(A+B)=P(A)+P(B) \tag{2}$$

又称可加性.

对立事件概率等式

$$P(\overline{A})=1-P(A) \tag{3}$$

是可加性的特例, 十分常用, 尚无通用简称. 本来 "逆概率公式" 是很适合的称呼, 可惜这个名称历史上早已用于称呼贝叶斯公式.

一般形式:

$$P(A_1\cup A_2\cup\cdots\cup A_n)=\sum_i P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)$$
$$-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n) \tag{4}$$

方便形式: 当 $A_1, A_2, \cdots, A_n$ 相互独立,

$$P(A_1\cup A_2\cup\cdots\cup A_n)=1-\prod_{i=1}^{n}\left[1-P(A_i)\right] \tag{5}$$

二、减法公式

$$P(A-B)=P(A)-P(AB) \tag{6}$$

特例: 当 $A\supset B$ 时,

$$P(A-B)=P(A)-P(B) \tag{7}$$

减法公式之称呼尚不通用, 在《概率论基础》的正文中也未出现, 有人用来指 (7) 式.

对立事件概率等式 (3) 也可以看作减法公式的特例.

三、乘法公式

$$P(AB)=P(A)P(B|A) \tag{8}$$

特例: 当 A 与 B 相互独立时,

$$P(AB) = P(A)P(B) \tag{9}$$

推广形式:

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{10}$$

一般形式:

$$P(A_1A_2\cdots A_n)=\sum_i P(A_i)-\sum_{i<j}P(A_i\cup A_j)+\sum_{i<j<k}P(A_i\cup A_j\cup A_k)$$

$$-\cdots+(-1)^{n-1}P(A_1\cup A_2\cup\cdots\cup A_n) \tag{11}$$

本式与(4)对偶, 但不常用,《概率论基础》仅在习题四的 1 题 (4) 中以示性函数形式给出提示.

四、全概率公式

$$P(B) = \sum_i P(A_i)P(B|A_i) \tag{12}$$

是从解题中总结出来的公式, 体现 "饭要一口一口吃". 它与下面贝叶斯公式成立的条件都是 $\sum_i A_iB = B$, 通常是 $\sum_i A_i = \Omega$.

全概率公式的使用一般都很自然, 通常不是冥思苦想的结果, 而是立即想到的第一选择.

五、贝叶斯公式

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)} \tag{13}$$

这个公式由贝叶斯作为推理公式提出, 近代发展 (例如在统计学与决策科学中) 应用广泛. 其重要性在概率论教本中实难充分体现.

§ 2.3　伯努利试验与直线上的随机游动

由雅科布·伯努利引进的重复独立试验模型是概率论发展史上最重要的模型, 概率论的许多重要结果都在这个模型中首先发

现. 从本节起, 现代概率论的序幕被拉开.

本节的重要论题有:

(1) 定义了 n 重伯努利试验;

(2) 对 n 重伯努利试验给出了概率空间 $(\Omega, \mathscr{F}, P)$;

(3) 导出伯努利, 二项, 几何, 帕斯卡等 4 个重要分布;

(4) 给分赌注问题以完满的解答, 参看教学札记之八 “分赌注问题”;

(5) 引入概率论中第一个动态模型——随机游动, 预示了随机过程的研究, 参看教学札记之九 “随机游动三题”;

(6) 解决了经典的赌徒输光问题;

(7) 把一维研究推广到多维, 并导出多项分布;

(8) 为讨论随机游动的常返性作了准备.

§ 2.4　二项分布与泊松分布

本节对二项分布作了仔细的研究, 带有示范性, 也留下许多伏笔. 泊松分布作为二项分布的逼近登场, 以泊松过程作结.

对二项分布的讨论从计算开始. 在当今, 一只科学计算器便应付裕如的计算, 在古代则是繁重的任务, 为解决这类难题, 许多创意由此产生, 不少理论因它建立. 通过对性质和图形的初步考察, 相信多数读者对二项分布已有一个整体的概念.

对二项分布的较深入讨论是结合一些应用展开的. 血清试验中的假设检验和抽样验收中的 (n, c) 方案涉及很深的统计问题, 四大应用实例显示了二项分布在应用中的重要性, 也提出了不可回避的计算问题, 将导致富有成果的研究.

历史上, 泊松分布正是作为小概率 p 下二项分布的近似计算公式引入的, 谁能料到现在它在概率论中的地位远在二项分布之上. 这里给出的泊松定理及证明大体保持历史原貌.

泊松分布的特质只有通过泊松过程才能显现.《概率论基础》把原属随机过程的内容写入基础概率论有点大胆, 但效果不错, 原

因之一是利用了辛钦对呼叫流的精细的分析研究成果 (见参考书目[8]), 使用柯西方程也有微功. 关于后者见教学札记之十一 "柯西方程与分布刻画".

特别指出, 研究动态随机现象 (随机过程) 是当代概率论的主题, 本章中有两个例证: 离散时间的随机游动与连续时间的泊松过程.《概率论基础》都用初等方法作了处理.

对随机游动, 一维无限制场合, 给出时刻 n 处于状态 k 的概率. 二维对称场合, 给出经 $2n$ 步返回原点的概率. 有吸收壁场合, 则利用全概率公式导出吸收概率所满足的差分方程, 用递推法得出答案. 进一步信息参看教学札记之九 "随机游动三题".

在泊松过程, 利用全概率公式导出状态概率 $P_k(t)$ 所满足的常微分方程, 经求解得出答案——泊松分布.

上述二者预示着有很多模型和大量结果可类似建立与导出.

习题解答与评注

1. 把字母 S、T、A、T、I、S、T、I、C、S 分别写在一张卡片上, 充分混合后重新排列, 问正好得到顺序 STATISTICS 的概率是多少?

提示: 用推广的乘法公式.

答 $\frac{3}{10}\cdot\frac{3}{9}\cdot\frac{1}{8}\cdot\frac{2}{7}\cdot\frac{2}{6}\cdot\frac{2}{5}\cdot\frac{1}{4}\cdot\frac{1}{3}\cdot\frac{1}{2}.$

【评注】这类问题用推广的乘法公式立即写出答案, 也可利用概率的古典定义和重复排列计算.

2. 若 M 件产品中包含 m 件废品, 今在其中任取两件, 求: (1) 取出的两件中至少有一件是废品的概率; (2) 已知取出的两件中有一件是废品的条件下, 另一件也是废品的条件概率; (3) 已知两件中有一件不是废品的条件下, 另一件是废品的条件概率.

解 (1) 设 A={取出的两件中至少有一件是废品},

$$P(A)=\frac{\binom{m}{1}\binom{M-m}{1}+\binom{m}{2}}{\binom{M}{2}}=\frac{m(2M-m-1)}{M(M-1)}$$

(2) 设 $A=\{$取出的两件中至少有一件是废品$\}$, 再设 $B=\{$取出的两件都是废品$\}$, 可见 $B\subset A$, 且有

$$P(B)=\frac{\binom{m}{2}}{\binom{M}{2}}$$

所求概率为

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{P(B)}{P(A)}=\frac{\binom{m}{2}}{\binom{m}{1}\binom{M-m}{1}+\binom{m}{2}}$$

$$=\frac{m-1}{2M-m-1}$$

(3) 设 $C=\{$取出的两件中有一件不是废品$\}$, $D=\{$取出的两件中恰有一件废品$\}$, 则所求的概率为

$$P(D|C)=\frac{P(CD)}{P(C)}=\frac{P(D)}{P(C)}=\frac{\binom{m}{1}\binom{M-m}{1}}{\binom{M-m}{2}+\binom{m}{1}\binom{M-m}{1}}$$

$$=\frac{2m}{M+m-1}$$

【评注】古典概型中条件概率直接计算一例.

3. 甲袋中有 a 只白球, b 只黑球, 乙袋中有 α 只白球, β 只黑球, 某人从甲袋中任取两球投入乙袋, 然后在乙袋中任取两球, 问最后取出的两球全为白球的概率是多少?

提示: 用全概率公式.

答 $$\frac{\binom{a}{2}\binom{\alpha+2}{2}+\binom{a}{1}\binom{b}{1}\binom{\alpha+1}{2}+\binom{b}{2}\binom{\alpha}{2}}{\binom{a+b}{2}\binom{\alpha+\beta+2}{2}}$$

【评注】全概率公式应用之典型例一.

4. 设一个家庭中有 n 个小孩的概率为

$$p_n=\begin{cases}\alpha p^n & , \quad n\geqslant 1\\ 1-\dfrac{\alpha p}{1-p}, & n=0\end{cases}$$

这里 $0<p<1$, $0<\alpha<\dfrac{1-p}{p}$, 若认为生一个小孩为男孩或女孩是等可能的, 求证一个家庭有 $k\,(k\geqslant 1)$ 个男孩的概率为 $2\alpha p^k/(2-p)^{k+1}$.

证 设 A_n={一个家庭中有 n 个孩子}, $n=1,2,\cdots$, B_k={该家庭中有 k 个男孩}, $k\geqslant 1$, 由于假定生男孩或生女孩是等可能的, 所以有 $P(B_k|A_n)=\binom{n}{k}\left(\dfrac{1}{2}\right)^n$, 由全概率公式得

$$\begin{aligned}P(B_k)&=\sum_{n=k}^{\infty}P(A_n)P(B_k|A_n)=\sum_{n=k}^{\infty}\alpha p^n\binom{n}{k}\left(\frac{1}{2}\right)^n\\&=\alpha\sum_{i=0}^{\infty}\binom{k+i}{k}\left(\frac{p}{2}\right)^{k+i}=\alpha\left(\frac{p}{2}\right)^k\sum_{i=0}^{\infty}\binom{k+i}{i}\left(\frac{p}{2}\right)^i\\&=\alpha\left(\frac{p}{2}\right)^k\sum_{i=0}^{\infty}\binom{-k-1}{i}\left(-\frac{p}{2}\right)^i=\alpha\left(\frac{p}{2}\right)^k\left(1-\frac{p}{2}\right)^{-k-1}\\&=\frac{2\alpha p^k}{(2-p)^{k+1}}\end{aligned}$$

【评注】全概率公式应用之典型例二.

5. 在上题假定下: (1) 已知家庭中至少有一个男孩, 求此家庭至少有两个男孩的概率; (2) 已知家庭中没有女孩, 求正好有一个男孩的概率.

解 (1) 设 A={家庭中至少有一个男孩}, B={家庭中至少有两个男孩}, 则有 $B \subset A$, 且

$$P(A)=\sum_{k=1}^{\infty}\frac{2\alpha p^k}{(2-p)^{k+1}}=\frac{2\alpha}{2-p}\cdot\frac{p/(2-p)}{1-p/(2-p)}$$

$$=\frac{\alpha p}{(2-p)(1-p)}$$

$$P(B)=\sum_{k=2}^{\infty}\frac{2\alpha p^k}{(2-p)^{k+1}}=\frac{2\alpha}{2-p}\cdot\frac{p^2/(2-p)^2}{1-p/(2-p)}$$

$$=\frac{\alpha p^2}{(2-p)^2(1-p)}$$

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{P(B)}{P(A)}=\frac{p}{2-p}$$

(2) 设 C={家庭中无女孩}={家庭中无孩子, 或家庭中有 n 个孩子且都是男孩}, 又设 D_1={家庭中正好有一个男孩}={家庭中只有一个孩子且是个男孩}, 则 $D_1 \subset C$, 且

$$P(C)=1-\frac{\alpha p}{1-p}+\sum_{n=1}^{\infty}\alpha p^n\left(\frac{1}{2}\right)^n$$

$$=1-\frac{\alpha p}{1-p}+\frac{\alpha p/2}{1-p/2}=1-\frac{\alpha p}{1-p}+\frac{\alpha p}{2-p}$$

$$=\frac{2-3p-\alpha p+p^2}{(1-p)(2-p)}$$

$$P(D_1)=\alpha p\cdot\frac{1}{2}=\frac{1}{2}\alpha p$$

所求概率为

$$P(D_1|C)=\frac{P(D_1C)}{P(C)}=\frac{P(D_1)}{P(C)}=\frac{\alpha p}{2}\cdot\frac{(1-p)(2-p)}{2-3p-\alpha p+p^2}$$

$$=\frac{\alpha p(1-p)(2-p)}{2(2-3p-\alpha p+p^2)}$$

【评注】以上两题给出一个人口生育的概率模型, 并用以训练全概率公式与条件概率的计算.

6. 已知产品中 96%是合格的, 现有一种简化的检查方法, 它把真正的合格品确认为合格品的概率为 0.98, 而误认废品为合格品的概率为 0.05, 求以简化法检查下为合格品的一个产品确实是合格品的概率.

提示: 用贝叶斯公式.

答 0.997 9.

【评注】贝叶斯公式最简应用举例.

7. 炮战中, 在距目标 250 米, 200 米, 150 米处射击的概率分别为 0.1, 0.7, 0.2, 而在各该处射击时命中目标的概率分别为 0.05, 0.1, 0.2, 现在已知目标被击毁, 求击毁目标的炮弹是由距目标 250 米处射出的概率.

提示: 用贝叶斯公式.

答 由贝叶斯公式可得所求概率为 0.043 5.

【评注】贝叶斯公式一般应用举例.

8. 飞机坠落在 A, B, C 三个区域之一, 营救部门判断其概率分别为 0.7, 0.2, 0.1; 用直升机搜索这些区域, 若有残骸, 被发现的概率分别为 0.3, 0.4, 0.5, 若已用直升机搜索过 A 区域及 B 区域, 没有发现残骸, 在这种情况下, 试计算飞机坠落在 C 区域的概率.

解 设 A={飞机坠落在 A 区域}, B={飞机坠落在 B 区域}, C={飞机坠落在 C 区域}, S={已搜索过 A, B 区域, 没有发现残骸}, 则

$$
\begin{aligned}
P(S) &= P(A)P(S|A) + P(B)P(S|B) + P(C)P(S|C) \\
&= 0.7 \times 0.7 + 0.2 \times 0.6 + 0.1 = 0.71
\end{aligned}
$$

所求概率为

$$P(C|S) = \frac{P(C)P(S|C)}{P(S)} = \frac{0.1}{0.71} = 0.14$$

【评注】把最新信息植入贝叶斯公式, 计算有关概率, 以助决策.

9. 选择题有 4 个答案, 只有一个是正确的. 不懂的学生从中随机选择. 假定一个学生懂与不懂的概率都是 1/2, 求答对的学生对该题确实懂的概率.

提示: 用贝叶斯公式.

答 0.8.

【评注】为选择题提供理论依据.

10. 甲袋中有 3 只黑球, 7 只白球; 乙袋中有 7 只黑球, 13 只白球; 丙袋中有 12 只黑球, 8 只白球. 先以 1 : 2 : 2 的概率选择甲、乙、丙中的一只袋子, 再从选中的袋子中先后摸出 2 球, 求: (1) 先摸到的是黑球的概率; (2) 已知后摸到的是白球, 求先摸到的是黑球的概率.

解 (1) 设 A_1、A_2、A_3 分别为选到甲袋、乙袋、丙袋的事件, B_1={先摸到的是黑球}, 则

$$P(B_1) = P(A_1)P(B_1|A_1) + P(A_2)P(B_1|A_2) + P(A_3)P(B_1|A_3)$$

$$= \frac{1}{5} \cdot \frac{3}{10} + \frac{2}{5} \cdot \frac{7}{20} + \frac{2}{5} \cdot \frac{12}{20} = 0.44$$

(2) 设 $\overline{B}_2$={后摸到的是白球}, 则

$$P(\overline{B}_2) = P(A_1)P(\overline{B}_2|A_1) + P(A_2)P(\overline{B}_2|A_2) + P(A_3)P(\overline{B}_2|A_3)$$

$$= \frac{1}{5} \cdot \frac{7}{10} + \frac{2}{5} \cdot \frac{13}{20} + \frac{2}{5} \cdot \frac{8}{20} = 0.56$$

这个答案也可由 (1) 利用对称性直接得到.

$$P(B_1\overline{B}_2) = P(A_1)P(B_1\overline{B}_2|A_1) + P(A_2)P(B_1\overline{B}_2|A_2) + P(A_3)P(B_1\overline{B}_2|A_3)$$

$$= \frac{1}{5} \cdot \frac{3}{10} \cdot \frac{7}{9} + \frac{2}{5} \cdot \frac{7}{20} \cdot \frac{13}{19} + \frac{2}{5} \cdot \frac{12}{20} \cdot \frac{8}{19} = 0.243\ 51$$

所以所求概率为

$$P(B_1|\overline{B}_2) = \frac{P(B_1\overline{B}_2)}{P(\overline{B}_2)} = \frac{0.243\ 51}{0.56} = 0.434\ 8$$

【评注】全概率公式与条件概率综合应用举例.

11. 甲、乙两人轮流射击, 先击中目标者获胜. 设甲、乙击中目标的概率分别为 p_1 及 p_2, 甲先射, 试求甲获胜的概率.

提示: 仔细分析, 从甲获胜这一事件入手.

答 $\dfrac{p_1}{p_1+p_2-p_1p_2}$.

【评注】独立性假定下的概率计算, 从事件分析中找出规律.

12. 飞机有三个不同的部分遭到射击, 在第一部分被击中一弹或第二部分被击中两弹, 或第三部分被击中三弹时, 飞机才能被击落, 其命中率与每一部分的面积成正比, 设三个部分的面积的百分比为 0.1, 0.2, 0.7. 若已击中两弹, 求击落飞机的概率.

提示: 分析两弹击中何处, 可击落飞机.

答 0.23.

【评注】还是从事件分析入手.

在初等概率论中, 事件分析和概率计算是训练的永恒主题.

13. 证明: 对于事件 A, B, 关系式

$$P^2(AB)+P^2(\overline{A}B)+P^2(A\overline{B})+P^2(\overline{A}\,\overline{B})=\frac{1}{4}$$

成立的充要条件为

$$P(A)=P(B)=\frac{1}{2},\quad P(AB)=\frac{1}{4}$$

证 柯西不等式

$$\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right)\geqslant\left(\sum_{i=1}^{n}a_ib_i\right)^2$$

等号成立的充分必要条件是

$$\frac{a_1}{b_1}=\frac{a_2}{b_2}=\cdots=\frac{a_n}{b_n}$$

由此立得

$$a^2+b^2+c^2+d^2=\frac{(a+b+c+d)^2}{4}$$

的充分必要条件是

$$a = b = c = d$$

进而若有

$$a + b + c + d = 1$$

则

$$a^2 + b^2 + c^2 + d^2 = \frac{1}{4}$$

的充要条件是

$$a = b = c = d = \frac{1}{4}$$

本题中

$$\begin{aligned} &P(AB) + P(\overline{A}B) + P(A\overline{B}) + P(\overline{A}\,\overline{B}) \\ &\quad = P\big((A + \overline{A})B\big) + P\big((A + \overline{A})\overline{B}\big) \\ &\quad = P(B) + P(\overline{B}) = 1 \end{aligned}$$

故由

$$P^2(AB) + P^2(\overline{A}B) + P^2(A\overline{B}) + P^2(\overline{A}\,\overline{B}) = \frac{1}{4}$$

推知

$$P(AB) = P(\overline{A}B) = P(A\overline{B}) = P(\overline{A}\,\overline{B}) = \frac{1}{4}$$

于是可得

$$P(A) = P(AB) + P(A\overline{B}) = \frac{1}{2}$$

$$P(B) = P(AB) + P(\overline{A}B) = \frac{1}{2}$$

反之, 由

$$P(A) = P(B) = \frac{1}{2} \text{ 及 } P(AB) = \frac{1}{4}$$

可推得

$$P(\overline{A}B) = P(B) - P(AB) = \frac{1}{4}$$

$$P(A\overline{B}) = P(A) - P(AB) = \frac{1}{4}$$

$$P(\overline{A}\,\overline{B}) = 1 - P(AB) - P(\overline{A}B) - P(A\overline{B}) = \frac{1}{4}$$

所以可得

$$P^2(AB) + P^2(\overline{A}B) + P^2(A\overline{B}) + P^2(\overline{A}\,\overline{B}) = \frac{1}{4}$$

【评注】从论证中可以看出恒成立不等式

$$P^2(AB) + P^2(\overline{A}B) + P^2(A\overline{B}) + P^2(\overline{A}\,\overline{B}) \geqslant \frac{1}{4}$$

而本题给出等号成立的充要条件. 这与习题一 45 题构成有趣的对照.

14. 若 A 与 B 独立, 证明 $\{\varnothing, A, \overline{A}, \Omega\}$ 中任何一个事件与 $\{\varnothing, B, \overline{B}, \Omega\}$ 中任何一个事件是相互独立的.

提示: 验证 16 个等式.

证 证明留给读者.

【评注】题意是, 事件 A 与事件 B 相互独立等价于由它们分别产生的 σ 域即 $\sigma(A)$ 与 $\sigma(B)$ 相互独立.

15. 若 $0 < P(B) < 1$, 试证:

(1) $P(A\,|\,B) = P(A\,|\,\overline{B})$;

(2) $P(A\,|\,B) + P(\overline{A}\,|\,\overline{B}) = 1$

均为 A 与 B 相互独立的充要条件.

证 (1) 若 A, B 相互独立, 则 $A, \overline{B}$ 也相互独立, 于是

$$P(A\,|\,B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

$$P(A\,|\,\overline{B}) = \frac{P(A\overline{B})}{P(\overline{B})} = \frac{P(A)P(\overline{B})}{P(\overline{B})} = P(A)$$

所以

$$P(A\,|\,B) = P(A\,|\,\overline{B})$$

反之, 若 $P(A\,|\,B)=P(A\,|\,\overline{B})$, 则

$$\frac{P(AB)}{P(B)}=\frac{P(A\overline{B})}{P(\overline{B})}=\frac{P(A)-P(AB)}{1-P(B)}$$

$$P(AB)-P(B)P(AB)=P(A)P(B)-P(B)P(AB)$$

所以

$$P(AB)=P(A)P(B)$$

故 A, B 相互独立;

(2) 若 A, B 相互独立, 则

$$P(A\,|\,B)+P(\overline{A}\,|\,\overline{B})=P(A)+P(\overline{A})=1$$

反之,

$$P(A\,|\,B)=1-P(\overline{A}\,|\,\overline{B})=P(A\,|\,\overline{B})$$

由 (1) 得 A, B 相互独立.

【评注】事件独立性有许多等价定义.

16. (费勒) 抽查一个家庭, 考察两个事件, A: 至多有一个女孩; B: 男女孩子都有. 假设男女的出生率都是 1/2, 试证: 对 3 个孩子之家, A 与 B 独立; 而对 4 个孩子之家, A 与 B 不独立.

证 对 3 个孩子之家:

$$\begin{aligned}P(A)&=P\{\text{至多有 1 个女孩}\}\\&=P\{\text{ 3 个男孩}\}+P\{\text{ 1 个女孩和 2 个男孩}\}\\&=\frac{1}{8}+3\times\frac{1}{8}=\frac{1}{2}\end{aligned}$$

$$\begin{aligned}P(B)&=P\{\text{男女孩子都有}\}\\&=P\{\text{1个女孩和2个男孩}\}+P\{\text{2个女孩和1个男孩}\}\\&=\frac{3}{8}+\frac{3}{8}=\frac{3}{4}\end{aligned}$$

$$P(AB)=P\{\text{ 1 个女孩和 2 个男孩}\}=3\times\frac{1}{8}=\frac{3}{8}$$

易得 $P(AB)=P(A)P(B)$, 所以 A, B 独立.

对 4 个孩子之家:

$$P(A) = P\{\text{4 个男孩}\} + P\{\text{1 个女孩和 3 个男孩}\}$$
$$= \frac{1}{16} + 4 \times \frac{1}{16} = \frac{5}{16}$$
$$P(B) = P\{\text{1个女孩和3个男孩}\} + P\{\text{2个女孩和2个男孩}\} + P\{\text{3 个女孩和 1 个男孩}\}$$
$$= \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{7}{8}$$
$$P(AB) = P\{\text{1 个女孩和 3 个男孩}\} = \frac{4}{16}$$

可见 $P(AB) \neq P(A)P(B)$, 故 A, B 不相互独立.

【评注】独立性在理论研究中应按定义验证, 但在应用中常用直观判断, 本题说明直观并不一定可靠.

有人说, 统计独立性并不是纯数学概念, 概率论中用 $P(AB) = P(A)P(B)$ 规范了它, 使之纳入体系而能处理, 这说法颇为深刻.

17. 事件 A, B, C 两两独立, $ABC = \varnothing$, $P(A) = P(B) = P(C)$, 且已知 $P(A \cup B \cup C) = \frac{9}{16}$, 试求 $P(A)$.

提示: 用一般加法公式.

答 $\frac{1}{4}$.

【评注】在第一章的计算题中加入独立性.

18. 设 A, B, C 三事件相互独立, 求证: (1) $A \cup B$, AB, $A - B$ 皆与 C 独立; (2) $\overline{A}, \overline{B}, \overline{C}$ 亦相互独立.

证 (1)
$$\begin{aligned} P((A \cup B)C) &= P(AC \cup BC) \\ &= P(AC) + P(BC) - P(ABC) \\ &= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C) \\ &= [P(A) + P(B) - P(AB)]P(C) \\ &= P(A \cup B)P(C) \end{aligned}$$

所以 $A\cup B$ 与 C 独立.

$$P\big((AB)C\big)=P(ABC)=P(A)P(B)P(C)=P(AB)P(C)$$

所以 AB 与 C 独立.

$$\begin{aligned}P\big((A-B)C\big)&=P(AC-BC)\\&=P(AC)-P(ABC)\\&=P(A)P(C)-P(AB)P(C)\\&=\big[P(A)-P(AB)\big]P(C)\\&=P(A-B)P(C)\end{aligned}$$

所以 $A-B$ 与 C 独立.

(2)
$$\begin{aligned}P(\overline{A}\,\overline{B})&=1-P(A\cup B)\\&=1-P(A)-P(B)+P(AB)\\&=1-P(A)-P(B)+P(A)P(B)\\&=\big[1-P(A)\big]\big[1-P(B)\big]\\&=P(\overline{A})P(\overline{B})\end{aligned}$$

类似可得

$$P(\overline{B}\,\overline{C})=P(\overline{B})P(\overline{C}),\quad P(\overline{C}\,\overline{A})=P(\overline{C})P(\overline{A})$$

又

$$\begin{aligned}P(\overline{A}\,\overline{B}\,\overline{C})&=1-P(A\cup B\cup C)\\&=1-P(A)-P(B)-P(C)+P(AB)\\&\quad+P(BC)+P(CA)-P(ABC)\\&=1-P(A)-P(B)-P(C)+P(A)P(B)\\&\quad+P(B)P(C)+P(C)P(A)-P(A)P(B)P(C)\\&=\big[1-P(A)\big]\big[1-P(B)\big]\big[1-P(C)\big]\\&=P(\overline{A})P(\overline{B})P(\overline{C})\end{aligned}$$

所以 $\overline{A},\overline{B},\overline{C}$ 相互独立.

【评注】两事件独立场合《概率论基础》第二章 §2 推论 1 及推论 2 之类结论在三事件独立场合的推广. 在证明中可以看出《概率论基础》第二章 §2 定义 2.2.2 中四个等式同时成立的必要性.

注意即使假定 A 与 C, B 与 C 独立, 甚至假定 A, B, C 两两独立, 仍然推不出 $A\cup B$ 与 C 独立; 另一方面, 若 A 与 C, B 与 C 独立, 又 A 与 B 不相容, 则 $A+B$ 与 C 独立. 对 $A-B$ 与 C 的独立, 两两独立性也不够, 除非 $A\supset B$. 这两者都来源于 AB 与 C 独立, 要求 $P(ABC)=P(AB)P(C)$, 这时需有《概率论基础》中 (2.2.4) 型的等式成立.

不难通过伯恩斯坦反例来加以佐证. 事实上在该例中

$$P\{A\cup B\}=\frac{3}{4},\quad P\{AB\}=\frac{1}{4},\quad P\{A-B\}=\frac{1}{4},\quad P(C)=\frac{1}{2}$$

$$P\{(A\cup B)C\}=\frac{1}{4},\quad P\{(AB)C\}=\frac{1}{4},\quad P\{(A-B)C\}=0$$

在 A, B, C 两两独立下, 本题结论并不成立.

*19. 证明: 事件 $A_1,A_2,\cdots,A_n$ 相互独立的充要条件是下列 2^n 个等式成立:

$$P(\hat{A}_1\hat{A}_2\cdots\hat{A}_n)=P(\hat{A}_1)P(\hat{A}_2)\cdots P(\hat{A}_n)$$

其中 $\hat{A}_i$ 取 A_i 或 $\overline{A}_i$.

证 必要性: 因为事件交的运算和数的乘法运算都满足交换律, 所以不妨设 $\hat{A}_1,\cdots,\hat{A}_n$ 中前 m 个为 $\overline{A}_i$, 后 $n-m$ 个为 A_i 的形式.

当 $m=0$, 成立

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)$$

当 $m=1$,

$$P(\overline{A}_1A_2\cdots A_n)=P(A_2\cdots A_n)-P(A_1A_2\cdots A_n)$$

$$= P(A_2)\cdots P(A_n) - P(A_1)P(A_2)\cdots P(A_n)$$

$$= P(\overline{A}_1)P(A_2)\cdots P(A_n)$$

设 $m = k$ 时有

$$P(\overline{A}_1\cdots\overline{A}_k A_{k+1}\cdots A_n) = P(\overline{A}_1)\cdots P(\overline{A}_k)P(A_{k+1})\cdots P(A_n)$$

则当 $m = k + 1$ 时

$$\begin{aligned}
&P(\overline{A}_1\cdots\overline{A}_k\,\overline{A}_{k+1}A_{k+2}\cdots A_n)\\
&= P(\overline{A}_1\cdots\overline{A}_k A_{k+2}\cdots A_n) - P(\overline{A}_1\cdots\overline{A}_k A_{k+1}A_{k+2}\cdots A_n)\\
&= P(\overline{A}_1)\cdots P(\overline{A}_k)P(A_{k+2})\cdots P(A_n)\\
&\quad - P(\overline{A}_1)\cdots P(\overline{A}_k)P(A_{k+1})P(A_{k+2})\cdots P(A_n)\\
&= P(\overline{A}_1)\cdots P(\overline{A}_k)\big(1 - P(A_{k+1})\big)P(A_{k+2})\cdots P(A_n)\\
&= P(\overline{A}_1)\cdots P(\overline{A}_k)P(\overline{A}_{k+1})P(A_{k+2})\cdots P(A_n)
\end{aligned}$$

从而有欲证的 2^n 个等式成立:

$$P(\hat{A}_1\hat{A}_2\cdots\hat{A}_n) = P(\hat{A}_1)P(\hat{A}_2)\cdots P(\hat{A}_n)$$

充分性: 直接可得

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2)\cdots P(A_n)$$

对 $n-1$ 个事件的交, 有

$$\begin{aligned}
&P(A_1\cdots A_{i-1}A_{i+1}\cdots A_n)\\
&= P(A_1\cdots A_{i-1}A_iA_{i+1}\cdots A_n) + P(A_1\cdots A_{i-1}\overline{A}_iA_{i+1}\cdots A_n)\\
&= P(A_1)\cdots P(A_{i-1})P(A_i)P(A_{i+1})\cdots P(A_n)\\
&\quad + P(A_1)\cdots P(A_{i-1})P(\overline{A}_i)P(A_{i+1})\cdots P(A_n)\\
&= P(A_1)\cdots P(A_{i-1})P(A_{i+1})\cdots P(A_n)
\end{aligned}$$

由归纳推理可知独立性定义要求的 $2^n - n - 1$ 个等式均可成立, 所以 $A_1, A_2, \cdots, A_n$ 相互独立.

【评注】考虑用这 2^n 个等式作为 $A_1, A_2, \cdots, A_n$ 相互独立定义的优点与缺点.

20. 三个工作小组独立对某个密码进行破译, 如果他们成功的概率分别为 0.4, 0.5, 0.7, 试求该密码被成功破译的概率.

提示: 用《概率论基础》公式 (2.2.6).

答 0.91.

【评注】事件并的概率计算在独立场合不再困难.

21. 设 $A_1, A_2, \cdots, A_n$ 相互独立, 而 $P(A_k) = p_k$, 试求: (1) 所有事件全不发生的概率; (2) 诸事件中至少发生其一的概率; (3) 恰好发生其一的概率.

提示: 上题做法的公式化与推广

答 (1) $\prod\limits_{k=1}^{n}(1-p_k)$;

(2) $1-\prod\limits_{k=1}^{n}(1-p_k)$;

(3) $\sum\limits_{j=1}^{n} p_j \prod\limits_{k\neq j}^{n}(1-p_k)$.

【评注】相互独立假定下的典型计算题.

22. 当元件 K 或者元件 K_1 及 K_2 都发生故障时电路断开, 元件 K 发生故障的概率等于 0.3, 而元件 K_1, K_2 发生故障的概率各为 0.2, 求电路断开的概率.

提示: 画出简图, 分清并联或串联线路, 各选适用算式, 再综合而成.

答 0.328.

【评注】在这类常见的计算题中, 只选最简单的一个, 是希望学生能举一反三.

23. 说明 "重复独立试验中, 小概率事件必然发生" 的确切意思.

答 以 A_k 记事件{小概率事件 A 在重复独立试验的第 k 次

出现}, 其发生的概率记为 $P(A_k)=\varepsilon, 0<\varepsilon<1$. 由 21 题之 (2), 在 n 次独立试验中, A 至少发生一次的概率为 $1-(1-\varepsilon)^n \longrightarrow 1(n\to\infty)$, 这说明当独立试验次数无限增加时, 即使是小概率事件, 它至少发生一次的概率无限接近于 1, 从而可以认为它是必然会发生的.

【评注】为经常乱穿马路, 爱在不洁小摊进食者戒.

24. 甲、乙、丙三人进行某项比赛, 若三人胜每局的概率相等, 比赛规定先胜三局者为整场比赛的优胜者, 若甲胜了第一、三局, 乙胜了第二局, 问丙成为整场比赛优胜者的概率是多少?

提示: 仔细分析事件, 小心计算概率.

答 $\dfrac{2}{27}$.

【评注】解题关键仍在于对事件进行细致分析.

25. 设实验室器皿中产生甲类细菌与乙类细菌的机会是相同的, 若某次发现产生了 $2n$ 个细菌, 求 (1) 至少有一个甲类细菌的概率; (2) 甲、乙两类细菌各占其半的概率.

提示: 产生的甲类细菌的个数可视作二项分布.

答 (1) $1-\left(\dfrac{1}{2}\right)^{2n}$; (2) $\dbinom{2n}{n}\left(\dfrac{1}{2}\right)^{2n}$.

【评注】进入伯努利概型, 二项分布.

26. 掷硬币出现正面的概率为 p, 掷了 n 次, 求下列概率: (1) 至少出现一次正面; (2) 至少出现两次正面.

提示: 用二项分布.

答 (1) $1-(1-p)^n$; (2) $1-(1-p)^n-np(1-p)^{n-1}$.

【评注】$P(A)=1-P(\overline{A})$ 应用之一例.

27. 甲、乙均有 n 个硬币, 全部掷完后分别计算掷出的正面数, 试求两人掷出的正面数相等的概率.

提示: 二人掷出的正面数都服从二项分布且独立.

解　利用二项分布, 得

$$p=\sum_{k=0}^{n}\binom{n}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{n-k}\cdot\binom{n}{k}\left(\frac{1}{2}\right)^{k}\left(\frac{1}{2}\right)^{n-k}$$

$$=\left(\frac{1}{2}\right)^{2n}\sum_{k=0}^{n}\binom{n}{k}^{2}$$

$$=\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}$$

【评注】这也是对称随机游动, 经 $2n$ 步返回原点的概率.

28. 在伯努利试验中, 事件 A 出现的概率为 p, 求在 n 次独立试验中事件 A 出现奇数次的概率.

解　事件 A 出现偶数次的概率记为 a, 出现奇数次的概率记为 b, 则

$$a=\binom{n}{0}p^{0}(1-p)^{n}+\binom{n}{2}p^{2}(1-p)^{n-2}+\cdots$$

$$b=\binom{n}{1}p(1-p)^{n-1}+\binom{n}{3}p^{3}(1-p)^{n-3}+\cdots$$

则有 $a+b=(p+1-p)^{n}=1$, $a-b=(1-p-p)^{n}$, 于是得到事件 A 出现奇数次的概率为 $\dfrac{1}{2}-\dfrac{1}{2}(1-2p)^{n}$.

【评注】正确写出 b 的表达式即算答对, 但应当想到可以化简, 从而达到最终的简明答案.

*29. 在伯努利试验中, 若 A 出现的概率为 p, 试证在出现 m 次 $\overline{A}$ 之前出现 n 次 A 的概率, 即分赌注问题中甲最终取胜的概率, 可由 (2.3.13), (2.3.14), (2.3.15) 中的任一式子表出, 即它们是相等的.

证　由帕斯卡分布得

$$P_1=\sum_{k=0}^{m-1}\binom{n+k-1}{k}p^{n}q^{k}\tag{2.3.13}$$

或

$$P_2 = \sum_{k=n}^{\infty} \binom{m+k-1}{k} p^k q^m \tag{2.3.14}$$

利用二项分布得

$$P_3 = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k q^{n+m-1-k} \tag{2.3.15}$$

其中 $q = 1 - p$.

(1) 先证 $P_3 = P_1$, 即

$$\sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k q^{n+m-1-k} = \sum_{k=0}^{m-1} \binom{n+k-1}{k} p^n q^k$$

当 $m = 1$, 得 $p^n = p^n$, 等式成立;

当 $m = 2$,

$$\text{左边} = \binom{n+1}{n} p^n q + \binom{n+1}{n+1} p^{n+1} = p^n(nq+1)$$

$$\text{右边} = \binom{n-1}{0} p^n + \binom{n}{1} p^n q = p^n(1+nq)$$

等式也成立;

假设当 $m = r$ 时,

$$\sum_{k=n}^{n+r-1} \binom{n+r-1}{k} p^k q^{n+r-1-k} = \sum_{k=0}^{r-1} \binom{n+k-1}{k} p^n q^k \tag{*}$$

则当 $m = r + 1$时,

$$\sum_{k=n}^{n+r} \binom{n+r}{k} p^k q^{n+r-k}$$

$$= \sum_{k=n}^{n+r-1} \binom{n+r}{k} p^k q^{n+r-k} + \binom{n+r}{n+r} p^{n+r}$$

由熟知的等式 $\binom{n+r}{k}=\binom{n+r-1}{k}+\binom{n+r-1}{k-1}$

$$=q\sum_{k=n}^{n+r-1}\binom{n+r-1}{k}p^kq^{n+r-1-k}+\sum_{k=n}^{n+r-1}\binom{n+r-1}{k-1}p^kq^{n+r-k}$$

$$+\binom{n+r-1}{n+r-1}p^{n+r}$$

$$=q\sum_{k=n}^{n+r-1}\binom{n+r-1}{k}p^kq^{n+r-1-k}+\sum_{k=n}^{n+r}\binom{n+r-1}{k-1}p^kq^{n+r-k}$$

$$=q\sum_{k=n}^{n+r-1}\binom{n+r-1}{k}p^kq^{n+r-1-k}+\sum_{k=n+1}^{n+r}\binom{n+r-1}{k-1}p^kq^{n+r-k}$$

$$+\binom{n+r-1}{n-1}p^nq^r$$

第 2 项中令 $j=k-1$

$$=q\sum_{k=n}^{n+r-1}\binom{n+r-1}{k}p^kq^{n+r-1-k}$$

$$+p\sum_{j=n}^{n+r-1}\binom{n+r-1}{j}p^jq^{n+r-j-1}$$

$$+\binom{n+r-1}{n-1}p^nq^r$$

$$=\sum_{k=n}^{n+r-1}\binom{n+r-1}{k}p^kq^{n+r-1-k}+\binom{n+r-1}{n-1}p^nq^r$$

$$\overset{(*)}{=\!=}\sum_{k=0}^{r-1}\binom{n+k-1}{k}p^nq^k+\binom{n+r-1}{r}p^nq^r$$

$$=\sum_{k=0}^{r}\binom{n+k-1}{k}p^nq^k,$$

故由归纳推理得 $P_3=P_1$.

利用对称性可知

$$\sum_{k=m}^{n+m-1}\binom{n+m-1}{k}q^k p^{n+m-1-k}=\sum_{k=0}^{n-1}\binom{m+k-1}{k}q^m p^k \qquad (**)$$

(2) 再证 $P_2=P_3$, 即

$$\sum_{k=n}^{\infty}\binom{m+k-1}{k}p^k q^m=\sum_{k=n}^{n+m-1}\binom{n+m-1}{k}p^k q^{n+m-1-k}$$

利用帕斯卡分布性质知

$$\sum_{k=n}^{\infty}\binom{m+k-1}{k}p^k q^m=1-\sum_{k=0}^{n-1}\binom{m+k-1}{k}p^k q^m$$

$$\overset{(**)}{=\!=}1-\sum_{k=m}^{m+n-1}\binom{m+n-1}{k}p^{m+n-1-k}q^k$$

令 $j=m+n-1-k$

$$=1-\sum_{j=n-1}^{0}\binom{n+m-1}{j}p^j q^{m+n-1-j}$$

$$=\sum_{j=n}^{n+m-1}\binom{n+m-1}{j}p^j q^{n+m-1-j}$$

于是有三式彼此相等.

【评注】总算写完如此繁复的论证. 顺便指出亦可利用 $\dfrac{\mathrm{d}P_3}{\mathrm{d}q}=\dfrac{\mathrm{d}P_1}{\mathrm{d}q}$ 证明 $P_3=P_1$.

30. 袋中有 10 只黑球, 10 只白球, 从中将球一只只摸出, 求在第 9 次摸球时摸得第 3 只黑球的概率.

提示: 事件分析: 必须且只需前 8 次得 2 黑且第 9 次摸得黑球.

答 $\dfrac{\dbinom{10}{2}\dbinom{10}{6}}{\dbinom{20}{8}}\cdot\dfrac{8}{12}.$

【评注】请与帕斯卡分布的推导及结论对照.

31. 设有 N 个袋子, 每个袋子中装有 a 只黑球, b 只白球, 从第一袋中取出一球放入第二袋中, 然后从第二袋中取出一球放入第三袋中, 如此下去, 问从最后一个袋中取出一球而为黑球的概率是多少?

解 设 A_1={从第 1 袋中取出一球是黑球}, 则

$$P(A_1)=\frac{a}{a+b},\quad P(\overline{A}_1)=\frac{b}{a+b}$$

设 A_2={从第 1 袋中取出一球放入第 2 袋中, 再从第 2 袋中取出一球是黑球}, 则

$$P(A_2)=\frac{a}{a+b}\cdot\frac{a+1}{a+b+1}+\frac{b}{a+b}\cdot\frac{a}{a+b+1}=\frac{a}{a+b}$$

$$P(\overline{A}_2)=\frac{b}{a+b}$$

一般设 A_k={按上述规则取球放球, 并在第 k 袋中取出一球是黑球}, 且 $P(A_k)=\dfrac{a}{a+b}$, 则

$$P(A_{k+1})=P(A_k)P(A_{k+1}|A_k)+P(\overline{A}_k)P(A_{k+1}|\overline{A}_k)=\frac{a}{a+b}$$

由数学归纳法得

$$P(A_N)=\frac{a}{a+b}$$

【评注】开始进入全概率公式 - 差分方程题型.

32. 甲袋中有 $N-1$ 只白球和 1 只黑球, 乙袋中有 N 只白球, 每次从甲、乙两袋中分别取出一只球并交换放入另一袋中去. 这样经过了 n 次, 问黑球出现在甲袋中的概率是多少, 并讨论 $n\to\infty$ 时的情况.

解 设 A_n={经 n 次交换后, 黑球在甲袋中}, 则 $\overline{A}_n$={经 n 次交换后, 黑球在乙袋中}, 记 $p_n = P(A_n)$, $q_n = P(\overline{A}_n) = 1 - p_n$, 由全概率公式

$$\begin{aligned}p_n &= P(A_{n-1})P(A_n \mid A_{n-1}) + P(\overline{A}_{n-1})P(A_n \mid \overline{A}_{n-1}) \\ &= p_{n-1} \cdot \frac{N-1}{N} + q_{n-1} \cdot \frac{1}{N} \\ &= \frac{N-1}{N}p_{n-1} + \frac{1}{N}(1 - p_{n-1})\end{aligned}$$

得差分方程 $p_n = \dfrac{1}{N} + \dfrac{N-2}{N}p_{n-1} \quad (n \geqslant 1)$, 初始条件 $p_0 = 1$, 故

$$\begin{aligned}p_n &= \frac{1}{N} + \frac{N-2}{N}\left(\frac{1}{N} + \frac{N-2}{N}p_{n-2}\right) \\ &= \frac{1}{N} + \frac{1}{N}\frac{N-2}{N} + \cdots + \frac{1}{N}\left(\frac{N-2}{N}\right)^{n-1} + \left(\frac{N-2}{N}\right)^n \\ &= \frac{1}{N}\left[1 - \left(\frac{N-2}{N}\right)^n\right] \Big/ \left(1 - \frac{N-2}{N}\right) + \left(\frac{N-2}{N}\right)^n \\ &= \frac{1}{2} + \frac{1}{2}\left(\frac{N-2}{N}\right)^n\end{aligned}$$

当 $N \geqslant 2$ 且 $n \to \infty$ 时, $\lim\limits_{n\to\infty} p_n = \dfrac{1}{2}$.

【评注】① 应导出下面的递推公式, 以后可引用.

即对差分方程 $p_n = ap_{n-1} + b$, 其解为

$$p_n = a^n\left(p_0 - \frac{b}{1-a}\right) + \frac{b}{1-a} \quad (a \neq 1)$$

② 若用 28 题结果可得另一解法.

③ 亦可用全概率公式–差分方程法求解 28 题.

*33. 投硬币 n 回, 第一回出正面的概率为 c, 第二回后每次出现与前一次相同表面的概率为 p, 求第 n 回时出正面的概率, 并讨论 $n \to \infty$ 时的情况.

解 设 A_n={第 n 回时出正面}, 并记 $p_n = P(A_n)$, 由全概率

公式

$$
\begin{aligned}
p_n &= P(A_{n-1})P(A_n \mid A_{n-1}) + P(\overline{A}_{n-1})P(A_n \mid \overline{A}_{n-1}) \\
&= p_{n-1}\,p + (1-p_{n-1})\,(1-p) \\
&= (1-p) + (2p-1)p_{n-1}
\end{aligned}
$$

得差分方程 $p_n = (2p-1)p_{n-1} + (1-p)$, 初始条件 $p_1 = c$, 且 $\dfrac{1-p}{1-(2p-1)} = \dfrac{1}{2}$, 由 32 题评注①得

$$
p_n = (2p-1)^{n-1}\left(c - \frac{1}{2}\right) + \frac{1}{2}, \quad p \neq 1
$$

当 $0 < p < 1, n \to \infty$ 时, $\lim\limits_{n\to\infty} p_n = \dfrac{1}{2}$.

【评注】可讨论 $p=0, p=1, c=\dfrac{1}{2}$ 等特殊场合, 加深理解.

*34. 甲、乙两袋各装一只白球一只黑球, 从两袋中各取出一球相交换放入另一袋中, 这样进行了若干次. 以 p_n, q_n, r_n 分别记在第 n 次交换后甲袋中将包含两只白球、一只白球一只黑球、两只黑球的概率. 试导出 p_{n+1}, q_{n+1}, r_{n+1} 用 p_n, q_n, r_n 表出的关系式, 利用它们求 p_{n+1}, q_{n+1}, r_{n+1}的表达式, 并讨论当 $n \to \infty$ 时的情况.

解 由全概率公式

$$
p_{n+1} = 0 \cdot p_n + \frac{1}{4}q_n + 0 \cdot r_n = \frac{1}{4}q_n
$$

$$
q_{n+1} = 1 \cdot p_n + \frac{1}{2}q_n + 1 \cdot r_n = p_n + \frac{1}{2}q_n + r_n
$$

$$
r_{n+1} = 0 \cdot p_n + \frac{1}{4}q_n + 0 \cdot r_n = \frac{1}{4}q_n
$$

另有等式 $p_n + q_n + r_n = 1$, 从而推得

$$
q_{n+1} = 1 - \frac{1}{2}q_n
$$

初始条件 $p_0 = r_0 = 0$, $q_0 = 1$, 由 32 题评注①可得解

$$q_{n+1} = \frac{2}{3}\left[1 - \left(-\frac{1}{2}\right)^{n+2}\right]$$

并推得

$$p_{n+1} = r_{n+1} = \frac{1}{6}\left[1 - \left(-\frac{1}{2}\right)^{n+1}\right]$$

当 $n \to \infty$ 时,

$$\lim_{n\to\infty} p_{n+1} = \lim_{n\to\infty} r_{n+1} = \frac{1}{6}, \quad \lim_{n\to\infty} q_{n+1} = \frac{2}{3}$$

【评注】这是统计物理学颇为重要的一个模型.

35. 一个工厂出产的产品中废品率为 0.005, 任意取来 1 000 件, 试计算下面概率: (1) 其中至少有 2 件废品; (2) 其中不超过 5 件废品; (3) 能以 90% 的概率希望废品件数不超过多少?

提示: 用伯努利概型, 利用泊松逼近.

答 (1) 0.959 572; (2) 0.615 960; (3) 8.

【评注】二项分布用泊松逼近计算.

36. 试给出泊松试验的严格表述.

解 泊松试验指如下试验序列:

$$\begin{array}{lllll} E_{11} & & & & \\ E_{21}, & E_{22} & & & \\ E_{31}, & E_{32}, & E_{33} & & \\ \cdots\cdots & & & & \\ E_{n1}, & E_{n2}, & E_{n3}, & \cdots, & E_{nn} \\ \cdots\cdots & & & & \end{array}$$

同一序列的诸试验相互独立, 且每一试验中事件 A 出现的概率等于 p_n, 只与序列编号 n 有关.

泊松试验中, 若 $np_n \longrightarrow \lambda$, 则第 n 号序列中事件 A 的出现次数渐近服从泊松分布.

【评注】泊松试验是很不同于伯努利试验的另一个概率模型，它大大拓展了概率论的研究领域.《概率论基础》中关于电话呼叫流的讨论提供了它的一个一维模型，而在参考书目[8]《概率论与数理统计简明教程》p.46 – 48 中则提供了它的一个二维模型，可以参看.

37. 某厂长有 7 个顾问, 假定每个顾问贡献正确意见的百分比为 0.6, 现为某事可行与否而个别征求各顾问意见, 并按多数人的意见作出决策, 求作出正确决策的概率.

提示: 假定独立性, 用二项分布.

答 0.710 2.

【评注】二项分布的直接计算.

38. 一本 500 页的书, 共有 500 个错字, 每个字等可能地出现在每一页上, 试求在给定的一页上至少有 3 个错字的概率.

提示: 用泊松分布.

答 0.080 302.

【评注】泊松分布计算的实例之一.

39. 某商店中出售某种商品, 据历史记录分析, 每月销售量服从泊松分布, 参数为 7, 问在月初进货时要库存多少件此种商品, 才能以 0.999 的概率充分满足顾客的需要.

提示: 反查泊松分布表.

答 16.

【评注】泊松分布计算的实例之二.

40. 螺丝钉生产中废品率为 0.015, 问一盒应装多少只才能保证每盒中至少有 100 只好螺丝钉的概率不小于 80% (提示: 用泊松逼近, 设应装 $100 + k$ 只).

提示: 按原题提示, $\lambda = (100 + k) \times 0.015 \approx 1.5$.

答 102.

【评注】无此提示几乎不知从何下手，以后应牢记.

41. 某疫苗中所含细菌数服从泊松分布, 每 1 毫升中平均含

有一个细菌, 把这种疫苗放入 5 只试管中, 每试管放 2 mL, 试求: (1) 5 只试管中都有细菌的概率; (2) 至少有 3 只试管中有细菌的概率.

提示: 每只试管的细菌数服从 $\lambda=2$ 的泊松分布, 各试管间的细菌数独立.

答 (1) 0.483 3; (2) 0.98.

【评注】泊松分布结合二项分布.

42. 实验室器皿中产生甲、乙两类细菌的机会是相等的, 且产生 k 个细菌的概率为

$$p_k=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\quad k=0,1,2,\cdots$$

试求: (1) 产生了甲类细菌但没有乙类细菌的概率; (2) 在已知产生了细菌而且没有甲类细菌的条件下, 有 2 个乙类细菌的概率.

解 (1) 产生了 k 个细菌且 k 个细菌全是甲类细菌的概率为 $\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\left(\frac{1}{2}\right)^k$, 所以所求概率为

$$\sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}\left(\frac{1}{2}\right)^k=\mathrm{e}^{-\lambda}\left(\mathrm{e}^{\frac{\lambda}{2}}-1\right)$$

(2) 产生了细菌但没有甲类细菌的概率与 (1) 中的概率相同, 产生 2 个乙类细菌的概率为 $\frac{\lambda^2}{2!}\mathrm{e}^{-\lambda}\left(\frac{1}{2}\right)^2$, 故所求的条件概率为

$$\frac{\lambda^2}{2!}\mathrm{e}^{-\lambda}\left(\frac{1}{2}\right)^2\Big/\mathrm{e}^{-\lambda}\left(\mathrm{e}^{\frac{\lambda}{2}}-1\right)=\frac{\lambda^2}{8\left(\mathrm{e}^{\frac{\lambda}{2}}-1\right)}$$

【评注】结合了本章所讲的许多概念.

43. 若每条蚕的产卵数服从泊松分布, 参数为 λ, 而每个卵变为成虫的概率为 p, 且各卵是否变为成虫彼此独立, 求每蚕养活 k 只小蚕的概率.

解 以 A_n 记蚕产出 n 个卵, 以 B_k 记每条蚕养活 k 只小蚕, 依独立性假定, 若蚕产出 n 个卵, 则这 n 个卵中成活的成虫数服从

二项分布, 于是所求概率为

$$P(B_k) = \sum_{n=k}^{\infty} P(A_n)P(B_k|A_n) = \sum_{n=k}^{\infty} \frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}\binom{n}{k}p^k(1-p)^{n-k}$$

$$= \frac{p^k\mathrm{e}^{-\lambda}}{k!}\sum_{j=0}^{\infty}\frac{\lambda^{k+j}}{j!}(1-p)^j$$

$$= \frac{1}{k!}(\lambda p)^k\mathrm{e}^{-\lambda p}$$

【评注】这是复合泊松分布的特例, 有多种变体, 以后还要讨论.

44. 通过某交叉路口的汽车流可看作泊松过程, 若在一分钟内没有车的概率为 0.2, 求在 2 分钟内有多于一车的概率.

提示: 先求 λ 值, 后算概率.

答 0.831 2.

【评注】泊松分布常见计算题的一种. 知道某个概率就能求 λ 值, 知道 λ 就能算一切概率.

45. 若已知 $t=0$ 时, 某分子与另一分子碰撞, 又知对任何 $t \geqslant 0$ 和 $\Delta t > 0$, 若不管该分子在时刻 t 以前是否遭受碰撞, 在 $(t, t+\Delta t)$ 中遭到碰撞的概率等于 $\lambda\Delta t + o(\Delta t)$, 试求该分子在时刻 τ 还没有再受到碰撞的概率.

解 记 $P(t)$ 为在时刻 t 还没有再受到碰撞的概率, $P(\Delta t)$ 为在 $(t, t+\Delta t)$ 中未遭到碰撞的概率, 依题意有

$$P(t+\Delta t) = P(t)P(\Delta t) = P(t)\left[1 - \lambda\Delta t - o(\Delta t)\right]$$

$$\frac{P(t+\Delta t) - P(t)}{\Delta t} = -\lambda P(t) - \frac{P(t)o(\Delta t)}{\Delta t}$$

令 $\Delta t \to 0$, 得微分方程:

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = -\lambda P(t)$$

及初始条件 $P(0)=1$, 解该微分方程得

$$P(\tau) = \mathrm{e}^{-\lambda\tau}$$

【评注】显示分析方法在概率论中应用的一个小小例子.

*46. 利用概率论的想法证明下面恒等式:

$$\sum_{k=0}^{N}\binom{N+k}{k}\frac{1}{2^k}=2^N$$

证 (巴拿赫火柴盒问题) 数学家的左、右衣袋中各放有一盒装有 N 根火柴的火柴盒, 每次抽烟时任取一盒用一根, 在发现一盒用光时, 另一盒有 r 根的概率为:

$$\binom{2N-r}{N}2^{-2N+r}$$

易见

$$\sum_{r=0}^{N}\binom{2N-r}{N}2^{-2N+r}=1$$

上式两边乘以 2^N 并利用组合性质得

$$\sum_{r=0}^{N}\binom{2N-r}{N-r}2^{-(N-r)}=2^N$$

令 $k=N-r$, 则有

$$\sum_{k=N}^{0}\binom{N+k}{k}2^{-k}=2^N$$

即

$$\sum_{k=0}^{N}\binom{N+k}{k}\frac{1}{2^k}=2^N$$

【评注】用的是 "在发现一盒用光时, 另一盒用去 k 根的概率". 同类题有习题一 *32 题.

47. 某车间宣称自己产品的合格率超过 99%, 检验人员从该车间的 10 000 件产品中抽查了 100 件, 发现有两件次品, 能否据此断定该车间谎报合格率?

解 因产品量很大, 所以可认为抽查的 100 件中的次品数服从二项分布, 不妨设合格率为 0.99, 则 100 件中次品不少于 2 件的概率为 (利用泊松逼近)

$$\begin{aligned}p &= 1-(0.99)^{100}-100\times 0.01\times(0.99)^{99}\\ &\approx 1-\mathrm{e}^{-1}-\mathrm{e}^{-1}\\ &= 0.264\,2\end{aligned}$$

这个概率不算小, 所以不能据此推定该车间谎报合格率.

【评注】统计假设检验法之例, 可以与《概率论基础》第二章 §4 例 2 (血清的试验) 对照.

48. 产品验收方案规定: 在一批 20 件产品中, 抽取其中 4 件, 若发现 1 件或 0 件次品, 则接受此批产品. 如果一批 20 件产品中含有 5 件次品, 若依上述方案验收, 试求这批产品被接受的概率.

提示: 化为古典概率计算.

答 0.751 3.

【评注】$(n,\ c)$ 方案与 OC 曲线实例.

*49. 系统中每个元件正常工作的概率为 p, 有半数元件正常则系统可工作, 对什么 p 值, $2k+1$ 个元件的系统比 $2k-1$ 个元件的系统好?

解 设 p_{2k+1}, p_{2k-1} 分别为 $2k+1$ 个元件和 $2k-1$ 个元件的系统正常工作的概率, 又以 X 记前 $2k-1$ 个元件中工作正常的元件数, 则

$$\begin{aligned}p_{2k+1} &= P(X=k-1)P(\text{后 2 个元件都工作正常})\\ &\quad + P(X=k)P(\text{后 2 个元件中至少有 1 个工作正常})\\ &\quad + P(X\geqslant k+1)\\ &= P(X=k-1)\,p^2+P(X=k)\big(1-(1-p)^2\big)+P(X\geqslant k+1)\end{aligned}$$

$$p_{2k-1} = P(X=k)+P(X\geqslant k+1)$$

$$p_{2k+1}-p_{2k-1}$$

$$= P(X=k-1)p^2 - P(X=k)(1-p)^2$$

$$= \binom{2k-1}{k-1}p^{k-1}(1-p)^k p^2 - \binom{2k-1}{k}p^k(1-p)^{k-1}(1-p)^2$$

$$= \binom{2k-1}{k}p^k(1-p)^k(2p-1)$$

要使 $p_{2k+1} > p_{2k-1}$ 当且仅当 $\frac{1}{2} < p < 1$.

【评注】直观解释: 若以少数服从多数的表决方式作出决策, 则当大家作出正确选择的概率超过 $\frac{1}{2}$ 时, 参加的人数越多越好. 类似地, 在体育比赛中, 对高手来说比赛的局数越多越有利.

**50. 通过构造适当的概率模型证明: 从正整数中随机地选取两数, 此两数互素的概率等于 $\frac{6}{\pi^2}$.

证明 如果以递增的次序将素数记上足标:

$$p_1 = 2,\ p_2 = 3,\ p_3 = 5,\ \cdots,\ p_k,\ \cdots$$

并以 A_{p_i} 记所取的第一个数能被 p_i 整除这一事件, 则

$$P(A_{p_i}) = \frac{1}{p_i}$$

不难验证,

$$P(A_{p_i}A_{p_j}) = P(A_{p_i})P(A_{p_j})$$

即 A_{p_i} 与 A_{p_j} $(i \neq j)$ 独立, 这个结果可推广到多个互素数的场合. 同样以 B_{p_j} 记所取的另一个数能被 p_j 整除的事件, 则

$$P(B_{p_j}) = \frac{1}{p_j}$$

且 B_{p_i} 与 B_{p_j} $(i \neq j)$ 独立. 由选数的随机性, 又可知

$$P(A_{p_k}B_{p_l}) = P(A_{p_k})P(B_{p_l})$$

以 C 记随机取两数, 此两数互素的事件, 即两数没有公因子,

或两数不能同时被某一素数整除，也就是

$$C=\bigcap_{i=1}^{\infty}\left(\overline{A}_{p_i}\cup\overline{B}_{p_i}\right)$$

由本章第 18, 19 题可知，$\overline{A}_{p_i}\cup\overline{B}_{p_i}$, $i=1,2,\cdots$, 相互独立，所以有

$$P(C)=\prod_{i=1}^{\infty}P(\overline{A}_{p_i}\cup\overline{B}_{p_i})=\prod_{i=1}^{\infty}\left[1-P(A_{p_i}B_{p_i})\right]$$

$$=\prod_{i=1}^{\infty}\left[1-P(A_{p_i})P(B_{p_i})\right]=\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^2}\right)$$

下一步证明

$$\prod_{i=1}^{\infty}\frac{1}{1-\frac{1}{p_i^2}}=\sum_{n=1}^{\infty}\frac{1}{n^2}$$

由几何级数公式:

$$\frac{1}{1-\frac{1}{p_i^2}}=1+\frac{1}{p_i^2}+\frac{1}{\left(p_i^2\right)^2}+\cdots+\frac{1}{\left(p_i^n\right)^2}+\cdots$$

将不超过正整数 N 的所有素数对应的这样的级数相乘，则部分乘积就等于

$$\prod_{p_i\leqslant N}\frac{1}{1-\frac{1}{p_i^2}}=\sum_{n=1}^{\infty}{}'\frac{1}{n^2}=\sum_{n=1}^{N}\frac{1}{n^2}+\sum_{n=N+1}^{\infty}{}'\frac{1}{n^2}$$

这儿的一撇 “′” 表示累加号并没有管到所有的正整数，只管到那些由小于等于 N 的素数乘积构成的正整数，但显然不大于 N 的所有正整数是由不大于 N 的素数的乘积构成，故等式右边第一个累加是不标以一撇的，由此进一步有

$$0<\prod_{p_i\leqslant N}\frac{1}{1-\frac{1}{p_i^2}}-\sum_{n=1}^{N}\frac{1}{n^2}<\sum_{n=N+1}^{\infty}\frac{1}{n^2}$$

因级数 $\sum_{n=1}^{\infty}\frac{1}{n^2}$ 的收敛性, 所以令 $N\to\infty$ 即有

$$\prod_{i=1}^{\infty}\frac{1}{1-\frac{1}{p_i^2}}=\sum_{n=1}^{\infty}\frac{1}{n^2}$$

有多种途径可求得 $\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}$.

例如将函数 $f(x)=\begin{cases}0\ , & x\in[-1,0)\\ x^2\ , & x\in[0,1)\end{cases}$ 展开为傅里叶级数:

$$f(x)\sim\frac{1}{6}+\frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos n\pi x+$$

$$\frac{1}{\pi}\sum_{n=1}^{\infty}\left[\frac{(-1)^{n+1}}{n}+2\frac{(-1)^n-1}{n^3\pi^2}\right]\sin n\pi x$$

$$=\begin{cases}0\ , & x\in(-1,0]\\ \frac{1}{2}\ , & x=\pm1\\ x^2\ , & x\in(0,1)\end{cases}$$

令 $x=1$, 稍加整理就可得到 $\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}$.

于是得到所求事件的概率为 $\frac{6}{\pi^2}$.

【评注】算术密度因不满足可列可加性难于定义概率而著名, 这里特殊问题特殊处理, 聊备一格.

在教学札记之十 "算术密度" 中, 有所介绍.

习题总评

第二章习题的训练重点是熟练掌握概率的基本公式, 以及二项分布与泊松分布的计算. 加法公式, 乘法公式相对容易些, 全概

率公式用途极广, 形式多样, 是概率论中最常用的公式, 本章和以后各章从正文到习题到处都有它的身影. 全概率公式在习题中配置若干类型, 贝叶斯公式在基础概率论中基本上与全概率公式紧密相连.

二项分布与泊松分布通过习题介绍了它们的各种应用并指出一些计算的要点.

独立性是本章引进的最主要概念, 因此配置不少习题以助理解.

伯努利概型是概率论中最重要的模型, 关于它很少专门设题, 而是融入二项分布, 几何分布, 帕斯卡分布的问题中.

章后小议

与其他概率论教材相比, 本章是多出的一章, 其他教材的做法是讲过条件概率与事件独立性等概念之后, 直接进入以随机变量与分布函数为主题的第二章.

这种与众不同的做法是基于以下的考虑: 过早引入随机变量并把概率计算问题归结到对分布函数的计算容易引导学生忽视概率论本身的思想与方法而造成分析化倾向.《概率论基础》用更多篇幅处理离散场合和经典模型, 因而对伯努利试验的讨论明显加强, 对各种离散型分布的产生背景有明确的交待, 让读者感受更多概率直观和经受更多概率方法的训练, 这也减轻了下一章引进分布时因一大堆分布同时登场而对初学者造成的困惑.

用专节讲二项分布与泊松分布, 除强调它们的重要性与紧密联系外, 也把有关的应用集中处理, 并以应用实例的方式介绍了一些有实际背景的例题, 在历史上这类题目的计算困难导致中心极限定理的发现与研究,《概率论基础》大致显示这个发展过程, 这些例题的最后解决将在第五章实现.

通过引入原属随机过程论的泊松过程的结构分析, 除强调这个

结果的重要性,也为下章两列等待时间分布的对比打下基础.

使用全概率公式,利用独立性之类的假定,通过建立某类方程,处理动态模型是随机过程论中最有力的方法之一,本章处理随机游动及泊松过程就是它的范例.

教学札记之五

对立事件概率等式

一、引言

对立事件 A 与 $\overline{A}$ 定义为 $A\cap\overline{A}=\varnothing$, $A\cup\overline{A}=\Omega$, 也即 $\overline{A}=\Omega-A$.

由加法公式立刻得到

$$P(A)+P(\overline{A})=1 \tag{1}$$

上式可写成

$$P(A)=1-P(\overline{A}) \tag{2}$$

或

$$P(\overline{A})=1-P(A) \tag{3}$$

称为对立事件概率等式.

对立事件概率等式可以看作加法公式的特例及变形,也可看作减法公式

$$P(\overline{A})=P(\Omega)-P(A)=1-P(A) \tag{4}$$

的特例.

它是概率论中最简单的公式,也是最常用的公式之一,有明显的概率解释,不过至今尚无通用的简称.可惜逆概率公式已被用来称呼贝叶斯公式,否则将是适当的简称.

这个公式的简单不用解释,但是它的用法却不简单,值得一议.下面就围绕这个主题作些归纳.

二、用法要诀——避难就易

$P(A)$ 与 $P(\overline{A})$ 虽然用一个简单公式相连, 但它们计算的难度一般并不相同, 有时竟大相径庭, 这就为避难就易留下余地. 个别情况下, 求 $P(A)$ 十分困难繁复, 但求 $P(\overline{A})$ 却甚简便, 几乎成了唯一的通途. 至于具体情况下的处理, 大体是采用先试一个然后 "碰鼻子转弯" 的策略, 当然也有些熟路可寻.

《概率论基础》第一章 § 3 一导出对立事件概率等式 (1.3.13) 就举了两个例子来说明这一要诀, 一是用来解决德 · 梅尔问题, 另一用来解一个特殊的摸球问题.

下面是另一个例子, 取自美国数学竞赛试题.

题目: 一个随机数产生器只能自 $1, 2, \cdots, 9$ 这九个整数中选一, 并且选哪一个都是等可能的, 试确定 $n\,(n>1)$ 次选择之后所得的 n 个数的乘积能被 10 整除的概率.

若以 A 记所得 n 个数的乘积能被 10 整除这一事件, 则要求的是 $P(A)$. 对使 A 发生的有利场合分析发现情况相当复杂, 不易理出头绪. 因此便想到从 $P(\overline{A})$ 入手一试, 即要分析事件 $\overline{A}$, 它表示所得 n 个数的乘积不能被 10 整除. 为做到这点, 只要不出 "5" 就能做到. 进一步研究发现, 即使出现了 "5", 只要不出现偶数也构不成 10 的倍数. 因此定义

$$B = \{\text{所选 } n \text{ 个数中不含 } 5\}$$

$$C = \{\text{所选 } n \text{ 个数中含 } 5, \text{但不含 } 2, 4, 6, 8\}$$

这样一来

$$B \cap C = \varnothing, \quad \overline{A} = B + C$$

显然

$$P(B) = \frac{8^n}{9^n}$$

另一方面

$$P(C) = \frac{5^n - 4^n}{9^n}$$

这里的 5^n 表明 n 个数由 1, 3, 5, 7, 9 组成, 而 4^n 表明 n 个数由 1, 3, 7, 9 组成, 因此 5^n-4^n 是 n 个数中至少有一个 "5", 余下的由 1, 3, 7, 9 中选入的总数, 这时选出的数的乘积不能被 10 整除.

因此

$$P(\overline{A})=P(B)+P(C)=\frac{8^n+5^n-4^n}{9^n}$$

从而

$$P(A)=1-P(\overline{A})=\frac{9^n-8^n-5^n+4^n}{9^n}$$

三、配合分布函数计算

在引进随机变量 ξ 之后, 概率计算集中于分布函数 $F(x)$, 这时若记

$$A=\{\xi<x\}$$

则

$$\overline{A}=\{\xi\geqslant x\}$$

而

$$P(\overline{A})=P\{\xi\geqslant x\}=1-P\{\xi<x\}=1-F(x) \tag{5}$$

(5) 式的利用在两种场合很显著.

其一, 离散型分布尾端概率的计算, 例如泊松分布计算:

$$\sum_{k=3}^{\infty}p_k=1-p_0-p_1-p_2$$

这类计算在第二章习题中常见.

其二, 是在生存分析中用生存函数 $S(x)=P\{\xi\geqslant x\}$ 代替非负随机变量 ξ 的分布函数 $F(x)=P\{\xi<x\}$ 作为工作语言. 见习题三 9 题.

四、配合对偶原理作概率计算

对偶原理

$$\overline{\bigcup_{i=1}^{n}A_i}=\bigcap_{i=1}^{n}\overline{A}_i \tag{6}$$

$$\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A}_i \tag{7}$$

把概率论中的两个很重要的事件联系起来: 一个是 "至少发生其一", 另一个是 "一个也不发生", 因而相应地, 对立事件概率等式

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 - P\Big(\bigcap_{i=1}^{n} \overline{A}_i\Big) \tag{8}$$

$$P\Big(\bigcap_{i=1}^{n} A_i\Big) = 1 - P\Big(\bigcup_{i=1}^{n} \overline{A}_i\Big) \tag{9}$$

也成为很重要的计算公式.

到底是化并为交, 还是化交为并, 要视情况而定, 总是选易求的一个入手. 习题一提供许多例子, 值得读者作一个分析和归纳. 例如第 35 题, 第 36 题, 第 38 题都是把交化为并, 再利用一般加法公式而获得解决的.

五、独立性假定下的一般加法公式

当 $A_1, A_2, \cdots, A_n$ 相互独立时,

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 - \prod_{i=1}^{n} [1 - P(A_i)] \tag{10}$$

特别是当 $P(A_i) = p, \quad i = 1, 2, \cdots, n$ 时,

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 - (1-p)^n \tag{11}$$

是非常有用的公式.

《概率论基础》第二章 §2 特设一个小节来讨论它的应用. 要点有二:

(1) 独立等概率场合, 小概率事件在大量重复试验中必出现.

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = 1 - (1-p)^n \longrightarrow 1 \quad (0 < p < 1, \quad n \to \infty)$$

(2) 在可靠性理论中的应用, 串联对应于交运算, 并联对应于并运算. 在元件的可靠性独立的假定下, 各类可靠性计算都可以化为并、交运算并利用对立事件概率等式实现.

实际上, 同类的应用在第三章还在继续, 表现为极值分布的推导. 设 $X_1, X_2, \cdots, X_n$ 相互独立, 均服从 $F(x)$, 则

$$\begin{aligned} P\{\max(X_1, \cdots, X_n) < x\} &= P\{X_1 < x, \cdots, X_n < x\} \\ &= P\{X_1 < x\} \cdots P\{X_n < x\} \\ &= \left[F(x)\right]^n \end{aligned}$$

这里处理的是 $A_i = \{X_i < x\}$ 的交, 有独立性假设, 因此好办. 但当处理 $\min(X_1, \cdots, X_n)$ 的分布时, 就遇到困难, 因

$$\{\min(X_1, \cdots, X_n) < x\} = \left\{\bigcup_{i=1}^{n}(X_i < x)\right\}$$

所以 (10) 或 (11) 就起作用, 这时

$$\begin{aligned} P\{\min(X_1, \cdots, X_n) < x\} &= 1 - P\{\min(X_1, \cdots, X_n) \geqslant x\} \\ &= 1 - P\{X_1 \geqslant x, \cdots, X_n \geqslant x\} \\ &= 1 - P\{X_1 \geqslant x\} \cdots P\{X_n \geqslant x\} \\ &= 1 - \left[1 - F(x)\right]^n \end{aligned}$$

这不过是 (11) 的变形. 明白了这点, 对第三章 §3 中的有关推导就不会感到陌生了.

教学札记之六

全概率公式与贝叶斯公式

一、前言

对事件 $B, A_1, A_2, \cdots, A_n, \cdots$, 若成立 $B = \sum\limits_{i=1}^{\infty} A_i B$, 则有全概

率公式

$$P(B) = \sum_{i=1}^{\infty} P(A_i)P(B|A_i) \tag{1}$$

及贝叶斯公式

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum\limits_{j=1}^{\infty} P(A_j)P(B|A_j)}, \quad i = 1, 2, \cdots \tag{2}$$

常用的情况之一是 $\{A_i,\ i = 1, 2, \cdots\}$ 是 Ω 的一个分割 (亦称完备事件组), 即 $\sum\limits_{i=1}^{\infty} A_i = \Omega$, 这时显然成立 $B = \sum\limits_{i=1}^{\infty} A_i B$; 另外, 也有从 B 作不相容分割出发的. 实用上有限 n 与理论上可列和都常见. 有时干脆写成无限, 若事实上只有 n 项, 则理解成当 $i > n$ 时 $P(A_i) = 0$, 公式照常使用.

这是概率论最重要公式中的两个, 也是密切联系的两个.

二、全概率公式

化繁为简是解题的基本思路, 因此在 1657 年出版的惠更斯的《论赌博中的计算》中就有利用

$$P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A})$$

解题思想的不少例证. 可见全概率公式很早就被人们所认识. 事实上, 用全概率公式解题是一种很自然的想法, 而不是一种高超的技巧.

《概率论基础》正文与习题中用到全概率公式的地方很多, 通过这些反复进行的训练, 相信所有学生对它都会掌握得很好. 如果在学习之后, 同学们能对题型作些归纳整理, 肯定会有不少帮助.

在对随机现象作动态描述或处理 (随机过程) 中, 全概率公式甚至用得更多, 因为对各种发展可能作全面考察必然用条件化作分析, 这时全概率公式成为必要的工具. 在《概率论基础》中所见的例子是对随机游动用差分方程研究吸收概率 (赌徒输光问题).

历史上关于赌博延续时间的讨论构成第二个例子, 这些都在惠更斯与伯努利等早期的著作中出现, 并引起后继者的不断研究, 从而成为近代马尔可夫过程理论的一部分.

对连续时间过程,《概率论基础》用于导出泊松过程的方法也是典型, 通过全概率公式研究在 $[t,\ t+\Delta t]$ 内过程的可能发展情况来导出状态概率 $P_k(t)$ 所应满足的常微分方程, 从而得到它的完整表达式.

这类尝试在各种随机过程中进行, 最后终于由科尔莫戈罗夫在 1931 年写出了 "概率论中的解析方法" 一篇宏文, 进行总结发挥, 大大推动了随机过程理论的发展. 这一大类过程后由辛钦提议称为马尔可夫过程.

对全概率公式的应用也不断有所创新, 它被用于敏感问题的调查就是一例.

敏感问题 (吸毒, 赌博, 作弊) 的社会调查一直是个难题, 因为很难得到被调查人的配合. 最后统计学家终于设计出一种调查方案, 通过迂回, 得到答案. 该方法简介如下:

被调查人在无旁人在场的状态下面对一张问卷, 内有两个问题:

问题 A: 你的手机电话号码的末位数是偶数吗?

问题 B: 你是否作弊过?

后附有操作说明: 先从旁边的袋子中摸出一球, 若为黑球则回答问题 A, 若为白球则回答问题 B.

这个方案大大消除了被调查人的顾虑, 因而能得到真实的回答. 假定对 n 个人作调查, 得到 k 个 "是" 的回答, 那么统计学家就能利用全概率公式来推算作弊率 p.

记袋中的白球的比例为 π, 黑球的比例为 $1-\pi$, 则 $P(\text{白})=\pi$, $P(\text{黑})=1-\pi$, $P(\text{是}|\text{白})=p$, $P(\text{是}|\text{黑})=0.5$, 于是回答 "是" 的概率为

$$P=\pi p+(1-\pi)\times 0.5$$

从而

$$p = \frac{P - 0.5(1-\pi)}{\pi}$$

把频率 $\frac{k}{n}$ 作为概率 P 的估计代入, 则得

$$\widehat{p} = \frac{\frac{k}{n} - 0.5(1-\pi)}{\pi} \tag{3}$$

当 n 足够大时, 相信能得到较可靠的作弊率估计值.

显然, 在这个方案中为获得真实的信息作出了一定的牺牲. 回答问题 A 的人的答卷不但无用而且带来混淆. 还有, 若在袋中放入更多的白球能使信息的收集更加有效, 但这样一来又会增添被调查人的顾虑.

无论如何, 这个设计是聪明的.

三、贝叶斯公式

初看贝叶斯公式 (2) 不过是条件概率定义, 乘法公式和全概率公式的简单组合, 它表明

$$P(A_i|B) = CP(A_i)P(B|A_i), \quad i = 1, 2, \cdots, n$$

或

$$P(A_i|B) \propto P(A_i)P(B|A_i), \quad i = 1, 2, \cdots, n$$

事实上, 在一些应用中, 只关心哪个 i 使上述条件概率达到最大, 根本不必计算 C. 当然在另一些问题中又必须计算 C, 这时 C 的计算成为使用贝叶斯公式的一大难点, 导致大量研究, 例如上世纪末 MCMC 方法的大发展, 这也是动力之一.

不过, 现代概率论与统计学者都认识到贝叶斯公式实在含有深意, 贝叶斯其人实在了不起.

托马斯 · 贝叶斯 (1702—1761) 出生于伦敦一个新教家庭, 其父是牧师, 由家庭教师传授知识. 有学者推测, 贝叶斯的家庭教师之一可能是棣莫弗 (1667—1754). 棣莫弗是法国的新教教徒, 18

岁时因参加宗教骚乱被监禁两年, 获释后由法国移居气氛较好的伦敦, 一直以家庭教师为业来维持生活. 贝叶斯 12 岁时, 伯努利给莱布尼兹写信说 "贫困的棣莫弗在伦敦不得不以教数学为生". 棣莫弗是第一流分析学家和概率论学家, 受到牛顿等人的高度评价, 于 1697 年当选皇家学会会员, 而托马斯的父亲也是皇家学会会员, 托马斯在概率论的研究与棣莫弗的工作有关, 在论文中也曾多次提起后者, 但没有他们交往的直接证据.

托马斯 · 贝叶斯也做牧师, 一生未婚, 从事神学与数学研究, 但只写过两本小册子, 1731 年的一本关于神学, 1736 年的一本关于 "流数学", 1742 年被接纳为皇家学会会员, 1752 年退休, 1761 年逝世.

贝叶斯于逝世前 4 个月把两篇遗稿寄给好友 Richard Price (1723—1791), Price 是一位名人 (社会保障先驱者), 不负朋友所托, 最后这两篇文章都在皇家学会的刊物上发表. 一篇关于级数, 另一篇就是我们要介绍的 "论机会学说中一个问题". 该文于 1763 年 12 月 23 日在皇家学会上宣读, 刊于 1763 年的《哲学学报》, 200 年后在 1958 年的 Biometrika 中重刊.

贝叶斯论文名为 "An essay towards solving a problem in the doctrine of chance", 由 R.Price 注释出版. 中国读者可在 1992 年由中国统计出版社出版的普雷斯著《贝叶斯统计学》 中读到该文的中译文.

关于这篇著名的论文, 现作如下简介.

一开头作者就十分明确地提出论文要解决的问题.

问题: 给定一个未知事件发生和失败的次数, 求其在一次试验中发生的概率位于两个任意指定的概率度之间的机会.

若以 θ 记事件 A 在一次试验中出现的概率, 以 X 记 A 在 n 次独立试验中发生的次数, 而 $0 \leqslant a < b \leqslant 1$, 贝叶斯要求的是 $P\{a \leqslant \theta \leqslant b \mid X = x\}$.

在伯努利试验中, 已知 p, 求 X 的分布是概率论中研究的主

题, 这里则是已知 $X = x$, 要求 p 落在某一区间的概率, 事实上属于统计问题. 雅科布·伯努利已关心由试验结果求概率的问题, 后来丹尼尔·伯努利第一个提出这个问题, 现由贝叶斯完全明确表述并给予解答, 自然意义重大, 通常称为逆概率问题, 因此贝叶斯公式也常被称为逆概率公式. 在这些先驱者眼中, 概率统计本来就是一家.

为了解决这个问题, 贝叶斯先提出 7 个定义, 后建立 10 个命题.

其 7 个定义如下:

定义:

1. 如果几个事件其中一个发生时, 其他均不发生, 我们称几个事件是互不相容的.

2. 如果两个事件其中总有一个发生, 但两个不能同时发生, 我们称这两个事件是对立的.

3. 我们称一事件失败, 是指它没能发生, 或者它的对立事件发生.

4. 当一个事件已经发生, 或者已经失败, 我们称它被确定了.

5. 任一事件的概率, 等于应予以计算的依赖于该事件发生的预期值与该事件发生时所期望的值之比.

6. 我们所说的机会就是概率.

7. 事件是相互独立的, 如果其中任一事件的发生既不增加也不减少其他事件的概率.

这里的定义 6 强调问题中所说的机会就是概率. 而定义 5 则给出一个与众不同的概率定义. 据 Price 转述, 贝叶斯 “这么做的愿望是为了消除关于这个词的词义的所有争议, 因为在普通语言中, 针对它用于过去或未来的事件的不同情况, 持不同观点的人对它有不同的理解, 但是, 不论对它的理解有何不同, 都允许对依据过去的事情的真实性或任何将来事件的发生的预期作出估计, 越真实的事情越有价值, 或者说这事件越会发生. 于是, 他提出无论概

率这个词在什么情况下被使用, 都允许对它给出合适的测度, 因而不会管对这个词如何恰当理解."

其他定义与我们现在的理解几乎完全一致.

接下来, 贝叶斯提出 10 个命题及若干推论, 命题 3 给出乘法公式 $P(AB) = P(B)P(A|B)$, 命题 5 给出条件概率定义 $P(A|B) = \frac{P(AB)}{P(B)}$, 命题 7则导出二项分布, 论文的第一部分至此突然结束.

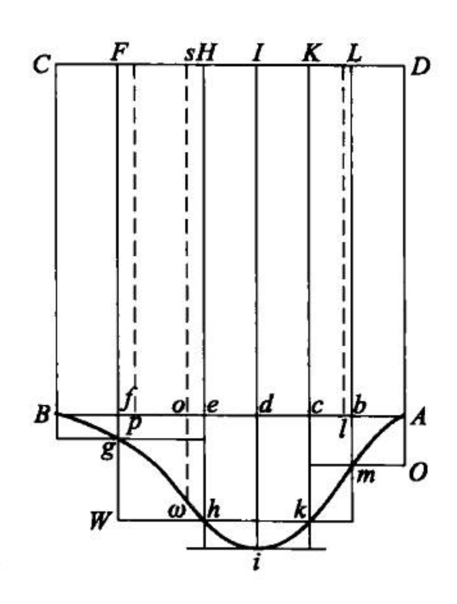


图 2-1

第二部分以如下台球桌模型开始.

假设:

1. 假定方球台或方平面 $ABCD$ 这样制成并水平放置, 使得将球 O 或球 W 中任意一个掷于其上时, 它落在其上的任意一个相等部分的概率都相同, 而且它必落在其上的某个部位.

2. 假定首先掷球 W, 通过它的落点画一条平行于 AD, 分别交 CD 和 AB 于 s 和 o 的线段 os, 然后掷球 O $p+q$ 次或 n 次, 如果在一次抛掷中它落在 AD 和 os 之间, 我们称事件 M 在一次试验中发生.

在这两个假定之后, 用几何方法证明了下面两条在我们看来是明显的引理.

引理 1 点 o 落在线段 AB 上的任意两点间的概率是这两点间的距离与线段 AB 之比.

引理 2 掷球 W, 画出 os, 那么在一次试验中事件 M 的概率为 $\frac{Ao}{AB}$.

接着又用繁复的几何方法证明了命题 8, 按现代记号即

命题 8 $P\{X=x,\ a\leqslant\theta\leqslant b\}=\int_a^b\binom{n}{x}\theta^x(1-\theta)^{n-x}\mathrm{d}\theta$

推论 $P\{X=x\}=\int_0^1\binom{n}{x}\theta^x(1-\theta)^{n-x}\mathrm{d}\theta=\frac{1}{n+1},$

$$x=0,1,2,\cdots,n,\quad n=1,2,\cdots$$

然后用命题 5 推出

命题 9

$$P\{a\leqslant\theta\leqslant b\,|\,X=x\}=\frac{(n+1)!}{x!(n-x)!}\int_a^b\theta^x(1-\theta)^{n-x}\mathrm{d}\theta$$

之后贝叶斯给出如下著名而引起巨大争议的“附注”:

附注 “……, 当我们在任何涉及某事件概率的试验进行前对该事件发生的概率还一无所知的情况下, 在此事件中应用这条原则是合适的. 似乎可以这样设想, 换句话说, 对于这样一个事件没有理由认为: 在一定数目的试验中, 它将更有可能发生某一次数而不是其他.”

这就是假定未知概率 θ 服从 $[0,1]$ 均匀分布, 后人称之为“同等无知”原则或“贝叶斯原则”.

这样就有了最后答案,

命题 10 $P\{a\leqslant\theta\leqslant b\,|\,X=x\}=\dfrac{\int_a^b\theta^x(1-\theta)^{n-x}\mathrm{d}\theta}{\int_0^1\theta^x(1-\theta)^{n-x}\mathrm{d}\theta}$

由于最后的答案包含了不完全 B 函数, 因此计算是一大问题, 这吸引了贝叶斯的许多注意力.

从现在的观点看, 贝叶斯的确提出了一整套归纳推理方案, 指明了解决问题的一种思路. 但与后来占主流地位的频率学派有很大分歧. 至少有如下几点:

(1) 把 θ 看作随机变量. 在频率学派看来,θ 虽未知,但是常数.

(2) 先验分布 $p(\theta)$ 的给定对最终答案关系重大, 如何给定?

(3) 对“同等无知”原则不认同, 例如对 θ“同等无知”与对 θ^2“同等无知”如何协调?

(4) 把概率看作信任程度的度量, 即主观概率.

古代的拉普拉斯, 近代的 Jeffreys, de Finetti, Wald, Savage, Lindley 等都对贝叶斯方法做出贡献, 使贝叶斯学派不断得到发展.

把概率作主观的信任程度的度量, 在许多学科 (例如金融学) 盛行, 与公理化体系也并不矛盾, 似乎不能简单排斥. 下面通过一个例子来说明.

《伊索寓言》中 "孩子与狼" 的故事和中国古代 "周幽王烽火戏诸侯" 的故事等信任度丧失的例子或可用贝叶斯公式来作定量分析.

某山村有一小孩, 每天上山放羊, 山里有狼出没. 有一天, 他兴致所至, 大喊 "狼来了, 狼来了". 山下村民闻言赶来才发觉受骗. 第二天仍然如此. 第三天真的有狼前来, 但不管孩子怎样叫也无人来救他了.

若以 $P(A)$ 记村民对该小孩的信任度, 假定为 0.8, 以 B_1 记该小孩有了第一次说谎, 设

$$P(B_1|A) = 0.1, \qquad P(B_1|\overline{A}) = 0.5$$

则

$$\begin{aligned} P(A|B_1) &= \frac{P(A)P(B_1|A)}{P(A)P(B_1|A) + P(\overline{A})P(B_1|\overline{A})} \\ &= \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.2 \times 0.5} = 0.444 \end{aligned}$$

这表明村民受第一次骗后, 对小孩的信任度由 0.8 降为 0.444.

第二次说谎后

$$P(A|B_2) = \frac{0.444 \times 0.1}{0.444 \times 0.1 + 0.556 \times 0.5} = 0.138$$

因此经两次说谎, 小孩在村民中的信任度已从 0.8 下降到 0.138, 难怪第三天无人上山搭救他.

教学札记之七

遗传学和概率论

一、前言

孟德尔 (Mendel, 1822—1884) 创始的遗传学理论对简单的概率模型的适用性提供了一个富有教益的例证.

遗传学规律是由奥地利修道士孟德尔发现的. 他从 1856 年在修道院的花园里种植豌豆, 开始他的 "豌豆杂交试验", 到 1864 年共进行了 8 年, 终于发现了前人未认识到的规律. 这些规律后来称为孟德尔定律, 是现代遗传学的基础.

孟德尔选用豌豆为试验植物, 从中选取了它的 7 个稳定的、易于区分的单位性状 (character) 作为观察分析的对象, 它们是种子的形状 (相对性状为圆粒和皱粒), 花的颜色 (红花和白花), 子叶的颜色 (黄色和绿色), 豆荚的形状 (膨大和缢缩), 豆荚的颜色 (绿色和黄色), 花的位置 (腋生和顶生), 茎的高度 (高茎和矮茎). 他在豌豆开花时进行品种之间授粉杂交的反复试验和精密统计分析, 终于在 34 个豌豆品种子孙后代的 2.8 万多棵植株里, 找到了遗传和变异的秘密.

孟德尔认为, 每一种性状都含有一种遗传因子, 生物的遗传性状都是由遗传因子决定的, 并提出了遗传学的两个基本定律: 分离定律 (law of segregation) 和独立分配定律 (law of independent assortment).

二、一对遗传因子的杂交试验

只有选择生物性状能够代代稳定相传的植株为亲本 (parents) 进行杂交, 才能确定子代的遗传基础. 豌豆是一种严格自花授粉的植物, 没有人为干预时, 多代自花授粉可形成纯系, 因此孟德尔选定了豌豆作为实验对象. 他从尚未成熟的花里除去雄蕊, 阻止了自

花授粉和自体授精, 到植株成熟的时候, 从具有相对性状的植株上采取花粉, 进行人工异花授粉, 这种杂交实验称为相互杂交 (孟德尔既做正交也做反交, 结果一致).

在杂交实验过程中, 孟德尔注意到由相互杂交所产生的子一代 (F_1) 的性状总是跟一个亲本的性状相似. 以圆粒种子和皱粒种子这一对相对性状为例说明. 圆粒种子的豌豆与皱粒种子的豌豆杂交得到的杂种子一代 F_1 全部为圆粒. 再让 F_1 自交得到子二代 F_2, 孟德尔从 253 棵植株上得到 7 324 粒 F_2 种子, 其中 5 474 粒为圆粒, 1 850 粒为皱粒, 二者的比例为 2.96 : 1. 孟德尔将 F_1 表现出来的亲本性状叫做显性性状 (dominant character), 把没有表现出来的亲本性状叫隐性性状 (recessive character), 本例中圆粒是显性性状而皱粒是隐性性状.

在豌豆的其他 6 对相对性状的杂交中, 孟德尔都得到了相同的试验结果: F_2 中两种性状的比例均接近于 3 : 1, 见下表:

性状的类别	亲本性状	F_1 性状	F_2 性状表现及数目	F_2比例
种子形状	圆粒 × 皱粒	圆粒	5 474 圆粒 1 850 皱粒	2.96:1
花的颜色	红花 × 白花	红色	705 红色 224 白色	3.15:1
子叶颜色	黄色 × 绿色	黄色	6 022 黄色 2 001 绿色	3.01:1
豆荚形状	膨大 × 缢缩	膨大	882 膨大 299 缢缩	2.95:1
豆荚颜色	绿色 × 黄色	绿色	428 绿色 152 黄色	2.82:1
花的位置	腋生 × 顶生	腋生	651 腋生 207 顶生	3.14:1
茎的高度	高茎 × 矮茎	高茎	787 高茎 277 矮茎	2.84:1

孟德尔为了解释这些结果, 提出下面的遗传因子分离假说.

(1) 生物的性状是由遗传因子控制的, 相对性状是由细胞中成对的遗传因子控制的.

(2) 遗传因子在体细胞中是成对的, 一个来自母本, 一个来自父本. 在形成配子时, 成对的遗传因子彼此分离, 并且各自分配到不同的配子中去, 每一个配子中, 只含有成对因子中的一个.

(3) 遗传因子在遗传的过程中始终保持为一个独立不变的遗传

单位, 杂种 F_1 体细胞内的遗传因子, 互不沾染.

(4) 成对遗传因子具有显隐性关系. 杂种一代表现一种表现型, 控制这种表现型的因子是显性因子, 相对应的是隐性因子.

(5) 杂种产生的不同类型的配子数目相等. 雌雄配子随机结合, 即不同类型的雌雄配子有同等的结合机会, 并且合子有同等的发育和存活的能力.

这样一来, 孟德尔对他的杂交试验结果进行了科学的解释. 通常将孟德尔的遗传因子用英文字母代表, 大写字母代表显性遗传因子, 对应的小写字母代表隐性遗传因子. 例如用 R 代表控制圆粒种子的显性遗传因子, r 代表控制皱粒种子的隐性遗传因子.

分离现象可用图 2–2 说明:

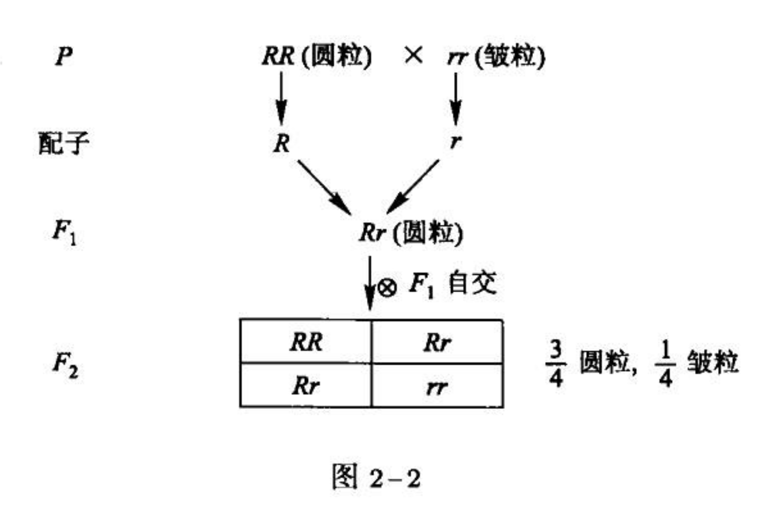


图 2–2

孟德尔的分离定律可表示如下: 成对的遗传因子的两个成员在形成配子时相互分离, 各自进入到不同的配子中去, 因此一半配子中携带成对遗传因子的一个成员, 而另一半配子携带成对遗传因子的另一个成员, 雌雄配子的结合是随机的. 一般情况下, 配子分离比是 1:1, F_2 遗传型分离比是 1:2:1, F_2 表现型分离比是 3:1.

孟德尔还设计了许多其他实验来验证上述结果, 都得到一致的结论, 从而建立了遗传因子的分离定律.

三、两对因子的杂交试验

根据孟德尔分离假说，一对遗传因子在形成配子时各自分离. 那么当两对或两对以上的遗传因子在形成配子时又有什么特征呢? 为此孟德尔又研究了两对遗传因子的遗传现象.

孟德尔仍以豌豆为实验材料，选取具有两对相对性状差异的纯系亲本进行杂交. 例如，用一个亲本是圆粒和黄色子叶的种子与另一个亲本是皱粒和绿色子叶的种子. 其子一代 F_1 都结圆粒和黄色子叶，表明圆粒和黄色子叶都是显性，这与 7 对性状分别进行研究的结果是一致的.

由 F_1 种子长成的植株 (共 15 株) 进行自交，得到 556 粒 F_2 种子，共有 4 种类型，其中两种类型和亲本相同，另两种类型为亲本性状的重新组合，而且存在着一定的比例关系，如图 2-3:

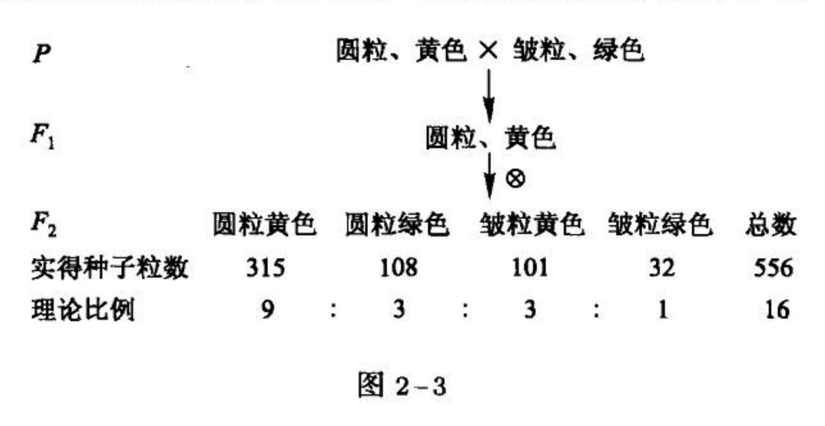


图 2-3

如果我们把种子形状和子叶颜色中的相对性状分别进行统计，可得如下结果:

种子类型		数　目	比例(%)
形状	圆粒	315+108=423	76.1
	皱粒	101+32=133	23.9
子叶	黄色	315+101=416	74.8
	绿色	108+32=140	25.2

每一对相对性状的分离比例都接近 3:1，符合分离规律.

如把两对性状联系在一起分析, F_2 表现型的分离比例是 9:3:3:1, 可通过下列表示法展示:

	3圆	:	1皱
×	3黄	:	1绿

9圆黄 : 3圆绿 : 3皱黄 : 1皱绿

图 2-4

孟德尔在实验的基础上总结出了独立分配定律, 亦称自由组合定律. 其要点是: 决定着相对性状的遗传因子在遗传的传递上有相对的独立性, 可以互不干涉, 在形成配子时, 可以互相随机地进行自由组合.

若以 R 和 r 分别代表种子的圆粒和皱粒, 以 Y 和 y 分别代表子叶的黄色和绿色. 那么纯系圆粒黄色种子亲本的遗传型为 $RRYY$, 纯系皱粒绿色种子亲本的遗传型为 $rryy$, 独立分配现象可用下图说明:

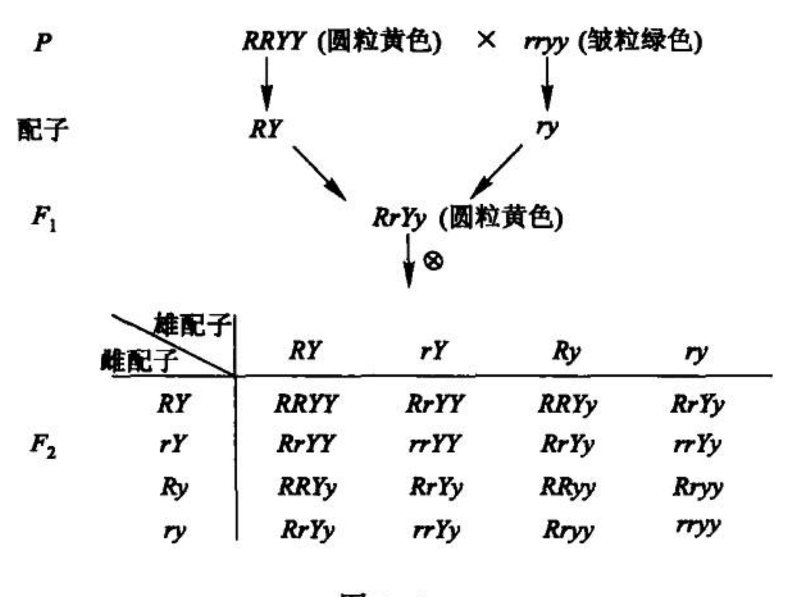


图 2-5

因此, 圆黄, 圆绿, 邹黄, 皱绿之比正好是 9:3:3:1.

孟德尔也设计了许多其他实验来验证遗传因子的独立分配定律,都得到一致的结论.

回归本题, 若用概率论语言, 则分离定律断言在 F_1 植株的花粉母细胞进行分裂时, 其所形成的雄配子总体中带有 R 和 r 的雄配子概率各为 $\frac{1}{2}$, 而独立分配定律则断言两对性状是相互独立的.

当然这些结论的成立是有条件的, 进一步的研究带来了孟德尔定律的补充与发展, 但孟德尔定律奠定了当代遗传学的坚实基础, 也宣布了概率统计成为研究生命现象的重要数学工具.

四、后语

1865 年孟德尔宣读了他的《植物杂交实验》论文并于次年正式发表, 但被长期埋没, 直至 1900 年另外三位不同国籍的科学家同时重新发现这些规律, 才重见天日. 从此孟德尔在生物领域具有里程碑性质的科学实验被载入史册. 1909 年丹麦遗传学家约翰逊正式命名孟德尔所称的 "遗传因子" 为 "基因" (Gene), 一直沿用至今.

下一个对遗传学做出重大贡献的是美国的摩尔根, 他通过对果蝇的研究, 发现连锁遗传, 建立了遗传学的第三条定律——连锁与互换定律, 创立了 "基因论", 并把抽象的基因概念落实到染色体上, 大大地发展了遗传学.

1953 年美国科学家沃森和英国科学家克里克发现了 DNA 的双螺旋结构, 对遗传学的发展做出重大贡献, 标志着分子遗传学时期的到来.

当代的一大盛事是人类基因组计划 (HGP) 的启动, 该计划已于 2003 年完成了人类基因组全部序列的测定, 正进入彻底阐明基因组编码的蛋白质的功能及 DNA 序列中所包含的遗传信息的生物功能的新阶段.

数学札记之八

分赌注问题

经过 150 多年对概率论发展早期史料的收集、整理、考订与研究, 目前大多数人已同意此前 (1837 年) 泊松 (1781—1840) 作出的论断: "由某广有交游者向严肃的詹森派信徒提出的一个博弈问题乃是概率演算的起源". 这个博弈问题就是分赌注问题, 古代一般称作点数问题.

一、分赌注问题的由来

分赌注问题大意如下: A, B 两个赌徒按某种方式下注赌博, 说定先胜 s 局者将赢得全部赌注, 但进行到 A 胜 s_1 局, B 胜 s_2 局 $(s_1 < s,\ s_2 < s)$ 时, 因故不得不中止, 试问如何分配这些赌注才公平合理?

这个分赌注问题大概相当古老, 因为至少在帕乔利 (Pacioli, 1445—1517) 出版于 1494 年的《算术、几何、比与比例集成》的专门论述 "稀奇问题" 的一节中已经出现, 其中 $s=6,\ s_1=5,\ s_2=2$. 这个问题知道的人也应该相当多, 因为上述书籍是当时流行的百科全书式数学教科书.

帕乔利把这个问题看作一个比例问题, 提出以 s_1 比 s_2 分赌注. 其论证思路是: 原定的先胜 s 局者为赢者的玩法至多在 $2(s-1)+1=2s-1$ 局内可完成, 因此赌注应按 $\dfrac{s_1}{2s-1}$ 比 $\dfrac{s_2}{2s-1}$ 分配. 在这个论证中既没有概率也没有组合.

这个答案受到当时最大的赌徒与数学家卡尔达诺 (Cardano, 1501—1576) 的反对, 他在发表于 1539 年的一篇文章中认为 "帕乔利按照已经赢得的局数成比例地分配赌金, 但没有考虑尚待每个赌徒去赢的局数." 卡尔达诺虽有概率想法也深通组合理论但对分赌注问题也提不出正确解法. 不过在当时的数学家中只有他意

识到分配规则不取决于 (s, s_1, s_2) 全体而只与未胜局数 $s-s_1$ 及 $s-s_2$ 有关.

另一个当时著名数学家塔塔利亚 (Tartaglia, 1499—1557) 在出版于 1556 年的《论数学与度量》中也批评帕乔利, 他举例说如果 A 先胜 10 点, B 未胜过, 难道赌注全归 A 所有? 并怀疑这个问题是否有数学答案, 他认为是一个司法的问题, 不过他也提出自己的答案: $s+s_1-s_2$ 比 $s-s_1+s_2$.

1558 年 Pereron, 1603 年 Forestani 都对此写过文章, 但都未提供正确答案.

二、帕斯卡与费马的通信

德·梅尔 (1607—1684) 就是那位 "广有交游者". 他与巴黎数学家有广泛联系, 也喜欢赌博, 后来还出入宫廷. 他向帕斯卡提出至少有两个问题, 一个掷骰子问题, 另一个就是分赌注问题, 正是这两个问题引起了帕斯卡与费马的通信.

帕斯卡 (1623—1662) 是历史上有名的 "数学神童", 8 岁后生活在巴黎. 父亲也是数学家, 他经常带儿子参加巴黎的各种科学聚会, 特别是每周一次的梅森家聚会. 梅森因数论中的梅森素数而在数学史上留名. 这种环境与经历让帕斯卡在 16 岁就写出《试论圆锥曲线》(1640), 20 岁发明手摇计算机, 对射影几何也有重要贡献, 成为该学科开拓者之一. 但他身体多病, 笃信神学, 身心颇受折磨.

费马 (1601—1665) 一生生活在法国南部的图卢兹, 在该地受教育, 成为律师, 后任议会顾问三十余年, 过着富足的生活. 30 多岁才研究数学, 全凭爱好、天赋和顽强精神, 成为数论学科的创始人, 并以大、小费马定理享誉数百年; 与笛卡尔平分解析几何发明权; 对于求曲线的切线和求平面曲线所围区域的面积也有诸多发现, 就是没有看出二者的联系而与微积分发明权失之交臂. 他从不发表文章, 许多证明也未写出. 他靠与当时知名数学家的通信互通信息. 他的文集是在他死后由他的儿子于 1679 年编辑出版的.

帕斯卡与费马关于概率论的通信共有 7 封:

1. 帕斯卡给费马第一封信(已丢失).

2. 费马给帕斯卡回信 (无日期).

3. 1654 年 7 月 29 日 (星期三) 帕斯卡给费马的第二封信.

4. 费马给帕斯卡信 (丢失).

5. 1654 年 8 月 24 日帕斯卡给费马的信.

6. 1654 年 8 月 29 日费马给帕斯卡的信.

7. 1654 年 9 月 25 日费马给帕斯卡的信.

从后面通信内容可以推测: 第一封信中帕斯卡向费马提出了掷骰子问题和分赌注问题.

费马在第二封信中对掷骰子的计算问题提出自己的解法并指出帕斯卡的一个错误.

第三封写于 7 月 29 日的长信是帕斯卡在病中完成的, 当时他很兴奋, 因为前一天他从 Carcavi 处收到费马的一封信, 谈到对分赌注问题的解法. 信中帕斯卡对费马的解法表示赞同, 承认自己的错误, 对分赌注问题提出自己的解法 (见下文), 也谈到德・梅尔曾向他和 Roberval 提出骰子问题.

第四封信虽丢失, 但可以肯定是费马对分赌注问题提出自己的解法, 这些解法在帕斯卡 8 月 24 日的回信即第五封信中有详细叙述.

第六封信中费马提到收到了 "论算术三角形" 拷贝. 从中可以看出帕斯卡在 "论算术三角形" 一文 (发表于 1665 年) 中发表的解法, 也列入这次通信讨论之中.

第七封信中费马提出 "Waiting time argument" 导出了现称为 "帕斯卡分布" 的表达式.

在这些通信中可以看出帕斯卡与费马都用组合方法正确解决了分赌注问题. 帕斯卡也用递推法, 并提出差分方程的显式解等同于组合解. 最后费马发现帕斯卡分布. 所有讨论都假定 A、B 两人的赌技相同, 即每人都有同等可能胜每一局, 这样才是古典概型.

两人都没有用“概率”一词，而代之以“值”，三年后惠更斯改称数学期望.

下面分别介绍二人的解法.

三、帕斯卡的解法

帕斯卡的初始解法出现在 7 月 29 日写的信中，对两人各投入 32 枚金币，以先得 3 分者为赢的场合作出详细解答. 原文如下:

> 下面给出在两个赌徒之间分配赌金的方法. 例如每人投放 32 枚金币作为赌金，并以先得 3 分为赢.
>
> 假设第一个人已得 2 分，另一个人只有 1 分. 他们掷下一次时，若第一个人赢了，他将得到全部 64 枚金币；若另一个人赢了，他们的比分是 2:2，如果在这种情况下分赌金的话，每人将拿回自己所下的赌金即 32 枚金币.
>
> 综上所述，第一个人如果赢了，64 枚金币将属于他；如果输了，32 枚金币将属于他. 假如他们不希望玩下去而要分赌金的话，第一个人应该说："我一定能得 32 枚，即使我下一轮输了，也应把它们给我. 至于另外的 32 枚金币，也许我得到它们也许你得到它们，机会是均等的. 所以，在我得 32 枚金币之后，再让我们均分另外的 32 枚吧." 这样，他将得到 48 枚金币，而另一个人只能得到 16 枚.
>
> 现在假定第一个人得 2 分而另一个人得 0 分，他们正在争夺下一分. 如果第一个人赢了，他将得到全部 64 枚金币. 如果另一个人赢了，注意他们将回到前面的情况，即第一个人有 2 分而另一个人有 1 分.
>
> 但我们已说明了在这种情况下，已有 2 分的人将得到 48 枚金币. 所以，如果他们不希望继续赌下去的话，这人应该说："如果我赢，我将得到全部 64 枚金币；如果我输了，48 枚金币将属于我. 所以请先把 48 枚金币给我，然

后再均分这剩下的 16 枚, 因为你我赢得它的机会是均等的." 于是, 他获得金币的数目为

$$48 \text{ 枚}+8 \text{ 枚} = 56 \text{ 枚}.$$

现在假定第一个人有 1 分而另一个人为 0 分. 先生请看, 如果他们再掷一次而第一个人赢了, 他与对手的比分将是 2 比 0. 根据前述理由, 56 枚金币将属于他. 如果他输了, 他们的比分成为 1 比 1, 32 枚金币将属于他. 所以他应该说: "如果你不打算赌下去, 就请把我原来的 32 枚金币给我, 再让我们把 56 枚金币的剩余部分均分. 56 减去 32 为 24, 让我们来平分这 24 枚金币吧. 你拿 12 枚我拿 12 枚, 再加上我原来的 32 枚, 我一共应得 44 枚金币."

从这个解法中可以看到几个要点: 其一是充分利用两人在每一局中胜负的等可能性; 其次是一种递推的思想. 这些在 "论算术三角形" 一文中都得到进一步阐述, 并结合组合公式的递推性推广到一般场合, 得到分赌注问题的正确解答公式.

"论算术三角形" 发表于 1665 年, 但是完成于 1654 年, 当时就有巴黎数学家看到此文, 而且费马在 8 月 29 日的通信中也提到收到该文的副本.

在 "论算术三角形" 中, 帕斯卡给出了一般结论: 若某人输时得赌金为 s, 而赢时得赌金 $s+t$, 则赌博中断时应得赌金 $s+\frac{t}{2}$. 从现在的观点看, 他用的是条件期望公式:

$$EX = \left[E(X|A) + E(X|\overline{A})\right]/2 \tag{1}$$

这里显然利用了等可能性. 至于递推法, 在一般场合, 若按现代术语, 以 $e(a,b)$ 记当 A, B 各缺 a, b 个胜局时 A 最终赢的概率. 这时可建立差分方程

$$e(a,b) = \left[e(a-1,b) + e(a,b-1)\right]/2 \tag{2}$$

边界条件则为

$$e(0,b) = 1, \quad e(a,0) = 0, \quad e(a,a) = \frac{1}{2}$$

利用算术三角形的性质及数学归纳法，帕斯卡最终得到

$$e(a,b)=\sum_{i=a}^{a+b-1}\binom{a+b-1}{i}\left(\frac{1}{2}\right)^{a+b-1}=\sum_{k=0}^{b-1}\binom{a+b-1}{k}\left(\frac{1}{2}\right)^{a+b-1} \tag{3}$$

并接近于建立递推公式

$$2\left[e(a,b)-e(a+1,b)\right]=\binom{a+b-1}{a}\left(\frac{1}{2}\right)^{a+b-1} \tag{4}$$

特例

$$a=1,\quad b=2,\quad e(a,b)=\frac{3}{4}$$

$$a=1,\quad b=3,\quad e(a,b)=\frac{7}{8}$$

$$a=2,\quad b=3,\quad e(a,b)=\frac{11}{16}$$

因此 A 得的赌金分别为

$$64\times\frac{3}{4}=48,\quad 64\times\frac{7}{8}=56,\quad 64\times\frac{11}{16}=44$$

与信中的结论一致.

帕斯卡利用组合理论，差分方程与递推法终于给分赌注问题以完满解答.

四、费马的解法

费马的解法主要通过帕斯卡在 8 月 24 日回信中复述而得知. 如两赌徒 A, B 离全胜所差局数分别为 a 与 b, 则最多再进行 $a+b-1$ 局就能决定胜负. 假定赌博继续进行，则共有 2^{a+b-1} 种等可能结果. 以 $a=2$, $b=3$ 为例，则有 $2^{2+3-1}=16$ 种等可能结果. 若以 A 表示 A 胜, B 表示 B 胜，则有以下 16 种情况:

$$\begin{array}{cccc} AAAA & AAAB & AABA & ABAA \\ BAAA & ABAB & ABBA & BABA \\ BBAA & BAAB & AABB & BBBA \\ BBAB & BABB & ABBB & BBBB \end{array}$$

其中对 A 有利的场合数为 11, 对 B 有利的场合数为 5, 因此 $e(a,b)=\frac{11}{16}$.

在回信中帕斯卡对费马的解法是折服的, 表示费马的方法很可靠, 并说已把费马的解法转告巴黎的一些数学家. 但有人持疑, 例如 Roberval 认为这一些皆在虚构中进行, 因而不可确信.

因为实际上的确赌博在一方达到胜利就结束, 很少要再进行到 $a+b-1$ 局的, Roberval 等数学家的疑虑是合理的, 这也推动费马再作进一步研究. 因此在费马于 9 月 25 日写给帕斯卡的信中, 特别指出在虚构的赌博中讨论点数问题并不影响赌金的分配结果. 例如对 $a=2, b=3$ 的情况, A 能够在再进行 2 局、3局和 4 局中取胜的方式为

$$AA;\quad ABA,\quad BAA;\quad ABBA,\quad BABA,\quad BBAA$$

每个排列都有 2 个 A 并且其中一个出现在最后位置. 在进行 2 局时, 有 2^2 种可能, 只有 AA 一种让 A 得胜; 在进行 3 局时有 2^3 种可能, 只有 ABA, BAA 2 种让 A 得胜; 在进行 4 局时有 2^4 种可能, 其中只有 3 种让 A 得胜, 故

$$e(2,3)=\frac{1}{2^2}+\frac{2}{2^3}+\frac{3}{2^4}=\frac{11}{16}$$

费马在信中也指出可将结果推广到更多局的场合.

最后费马用 “Waiting time argument” 找到

$$e(a,b)=\sum_{i=0}^{b-1}\binom{a-1+i}{i}\left(\frac{1}{2}\right)^{a+i} \tag{5}$$

这是《概率论基础》中 (2.3.13) 当 $p=q=\frac{1}{2}$ 时的特例. 该式用到现称的 “帕斯卡分布”.

通过上述讨论, 可见帕斯卡并未涉及这种分布, 而费马倒是达到它. 因此称这种分布为 “帕斯卡分布” 似有不妥.

在评述帕斯卡与费马的历史性突破时不要忘了当时的时代背景与科学水平. 在 1654 年代, 微积分尚未发明, 组合理论很不完

善. 须知排列还有待于这年出生的雅科布 · 伯努利来定名, 而组合符号 $\binom{n}{k}$ 要等欧拉 (1707 — 1783) 来引进.

五、惠更斯的贡献

惠更斯 (1629 — 1695) 是 17 世纪欧洲科学界的中心人物之一. 1655 年秋惠更斯为了得到安格斯基督教大学的法学博士学位到了法国并在巴黎住下. 他在那儿见到了许多有名的科学家. 他对 Mylon 和 Roberval 给他介绍的帕斯卡和费马所研究的新问题印象很深. 他还听到了分赌注问题. 但是没有人告诉他这些问题的解以及费马和帕斯卡的方法.

1655 年底惠更斯回到荷兰便独立研究这些问题并解决它们, 写成《论赌博中的计算》送其恩师 V. Schooten 审阅, 后者正在筹印《数学习题集》, 就建议他印刷发表, 并亲自替他译成拉丁文. 其后惠更斯通过 Carcavi 与费马联系上, 费马又向他提出 5 个新问题. 在 1657 年 3 月书稿最后一次校订时, 惠更斯将其论文增加为 14 个命题和 5 个问题, 该书出版于 1657 年 9 月, 以 1657 年 4 月 27 日惠更斯给 V. Schooten 的信作为序言, 而荷兰文版出版于 1660 年.

该书先给出一条公理和 14 个命题

公理 每个公平博弈的参与者愿意拿出经过计算的公平赌注冒险而不愿拿出更多的数额, 即赌徒愿意押的赌注不大于其获得赌金的数学期望.

命题 1 若得到 a 或 b 的机会相等, 则其值是 $\frac{a+b}{2}$.

命题 2 若得到 a, b 或 c 的机会相等, 则其值是 $\frac{a+b+c}{3}$.

命题 3 如果得到 p 次 a 和 q 次 b 的机会相等, 则其值是 $\frac{pa+qb}{p+q}$.

命题 4 假设两人一起赌博, 离全胜所差局数分别为 1, 2 时,

其赌注如何分配?

命题 5 假设两人一起赌博, 离全胜所差局数分别为 1, 3 时, 其赌注如何分配?

命题 6 假设两人一起赌博, 离全胜所差局数分别为 2, 3 时, 其赌注如何分配?

命题 7 假设两人一起赌博, 离全胜所差局数分别为 2, 4 时, 其赌注如何分配?

命题 8 假设三人一起赌博, 离全胜所差局数分别为 1, 1, 2 时, 其赌注如何分配?

命题 9 假设 n 个人一起赌博, 离全胜所差局数分别为 $r_1, r_2, \cdots, r_n$ 时, 其赌注如何分配?

命题 10－14 关于骰子问题, 从略.

惠更斯给这些命题一一证明. 重要的是在命题 3 中给出了分布

	a	b
	$\frac{p}{p+q}$	$\frac{q}{p+q}$

的数学期望

$$\mu = \frac{pa + qb}{p + q}$$

正是现代离散型数学期望的定义, 不过在荷兰文版中仍称 “值”.

对于点数问题, 惠更斯认识到与已胜局数无关, 而与离全胜所差的局数 a, b 相关, 而至多再进行的局数为 $a + b - 1$, 对 $e(a, b)$ 的求法惠更斯与帕斯卡基本一样.

惠更斯的书一出版就立即得到学术界的认可与重视. 在欧洲作为概率论的标准教材长达 50 年之久. 而帕斯卡与费马的通信是 1679 年才发表的, 因而惠更斯对早期概率论发展的贡献可与帕斯卡及费马并列.

六、雅科布·伯努利的贡献

雅科布·伯努利 (1654—1705) 在他死后由子侄整理出版的《猜度术》一书中, 第一部分即为惠更斯《论赌博中的计算》的重印与评注, 篇幅增到 5 倍, 可见下的功夫之深.

在关于点数问题中, 他评注说 "在计算分配额时, 只要注意即将进行的比赛, 不要管已经完成的比赛情况." 作为补充, 伯努利建立了一组提供二个赌徒的机会值和在各种条件下二个赌徒之间赌注分配率的曲线.

最重要的是他导出了对一般 p 的二项分布公式

$$\binom{n}{k}p^k q^{n-k}, \quad q = 1 - p$$

而二项分布是帕斯卡解与惠更斯解的主要组成部分, 这时已突破以等可能性为特征的古典概型, 适用于赌徒机会不等的情形.

七、蒙特莫特的贡献

蒙特莫特 (Montmort, 1678—1719) 是法国的数学家, 与伯努利家族有联系, 他在概率论方面的著作《Essay d'Analyse sur les Jeux de Hazard》(简称《随笔》) 第一版于 1708 年在巴黎出版, 该书介绍了伯努利《猜度术》的概要. 在该书的第三部分详细地讨论了帕斯卡与费马的通信, 特别是重新演算了 1654 年 8 月 24 日帕斯卡信中的全部题目, 也详细讨论了三人场合的分赌注问题.

该书出版后, 蒙特莫特送一本给约翰·伯努利, 并收到日期为 1710 年 3 月 17 日的回信, 在信中约翰指出点数问题对任意 p 值的解可由 $(p+q)^{a+b-1}$ 的展开式得到, 即

$$e(a, b) = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} p^i q^{a+b-1-i} \tag{6}$$

显然这是分赌注问题正确解的普遍表达式, 它通过把 (3) 中的 $\left(\frac{1}{2}\right)^{a+b-1}$ 用 $p^i q^{a+b-1-i}$ 代入而得到.

顺带指出, 同样的解被棣莫弗独立得到, 见 1711 年论文, 该文后扩充为《机会学说》(1718).

《随笔》的第二版于 1713 年出版, 蒙特莫特在点数问题部分加了三页, 他首先介绍约翰的解答, 其次他用二项式导出这个问题的解的一个新公式

$$e(a,b)=\sum_{i=0}^{b-1}\binom{a-1+i}{i}p^a q^i \tag{7}$$

这正是分赌注问题的另一个正确解的普遍表达式, 它通过把 (5) 中的 $\left(\frac{1}{2}\right)^{a+i}$ 用 $p^a q^i$ 代入而得到.

蒙特莫特的思路是: 在 n 次试验中成功 a 次的概率是 $\binom{n}{a}p^a q^{n-a}$, 可知要在第 $n+1$ 次试验中完成第 a 次成功的概率是

$$\binom{n+1}{a}p^a q^{n+1-a}-q\binom{n}{a}p^a q^{n-a}=\binom{n}{a-1}p^a q^{n+1-a} \tag{8}$$

对 $n+1=a+i$, $i=0,1,\cdots,b-1$ 求和即得 (7) 式.

蒙特莫特并指出在 (7) 中乘 q^i 以 $(q+p)^{b-1-i}$, 再把二项式展开经运算后可得到 (6), 从而证明了上述两个解相等.

关于蒙特莫特的上述思路解释如下: (8) 的第一式表示 A 到第 $n+1$次试验已胜 a 次; 第二式表示 A 在第 n 次试验就达到胜 a 次, 而第 $n+1$ 次试验失败; 第三式表示 A 到第 n 次试验共胜 $a-1$ 次而第 $n+1$ 次试验又胜. 概率等式 (8) 的成立不管从事件分析, 或者利用组合基本等式都可证明. A 要最终赢得赌金必须 $n+1=a+i$ 而 $i=0,1,\cdots,b-1$, 因此对其求和. 最右一式即为 (7), 表示在第 $a+i$ 次试验中 A 正好完成胜 a 次.

最后顺带提及, 在这本书中还刊载了尼古拉 · 伯努利 (Nicolas Bernoulli, 1687 — 1759) 的一封信, 在该信中已提出后来以 "圣彼得堡悖论" 闻名的问题. 参看教学札记之二十 "圣彼得堡悖论与期望效用函数".

八、分赌注问题的一般形式解

现代形式的分赌注问题之解是在伯努利概型中对任意 p $(0 < p < 1)$ 给出的, 在《概率论基础》第二章§3 中已作完整的介绍. 一种思路是沿着费马指引的设想赌到 $n+m-1$ 局, 必能结束, 这时利用二项分布求甲胜 $n, n+1, \cdots, n+m-1$ 局的概率, 此即为甲最终取胜的概率 (2.3.15), 相当于本文中的 (6), 因彼时乙达不到取胜 m 局.

另一思路是利用帕斯卡分布计算首次达到 r 次成功. 这时有两种具体方式: 其一考虑当甲再胜 n 局时乙的情况, 若乙的胜局为 $0, 1, \cdots, m-1$, 则甲已最终取胜, 这就是 (2.3.13) 所表达的, 相当于本文中的 (7); 其二是考虑当乙再胜 m 局时甲的情况, 若此时甲已胜 $n, n+1, \cdots$, 则甲早已捷足先登, 最终取胜拿走全部赌金, 就表示为 (2.3.14) 的无穷级数.

三个答案表面看来差别甚大, 假如不能证明它们是相等的, 那么分赌注问题就不算解决. 幸而经过繁复的计算, 我们已证明它们的确相等, 见习题 *29 题的解答, 这样我们也最终证实帕斯卡的话: "无论在图卢兹还是在巴黎, 真理是唯一的."

教学札记之九

随机游动三题

随机游动是概率论中研究的第一个随机过程, 它可以作为许多物理现象或社会现象的动态模型, 在理论上与应用中都十分重要. 下面只关心与《概率论基础》比较有关的三件事.

一、赌徒输光问题的一点历史注记

《概率论基础》中提到随机游动溯源于赌徒输光问题. 这一事实是公认的, 但各书讲法详略不一且颇有分歧, 现追索原始资料, 简介于下:

1655 年荷兰数学家惠更斯访问巴黎, 听其他数学家讲到了帕

斯卡和费马的新研究, 印象很深, 不过并不知道他们的解法. 年底回国后便独自研究这些问题并于 1657 年出版了《论赌博中的计算》. 这本书包括一篇简短的序言和 14 个命题. 在这些命题及其解答中惠更斯明确提出了数学期望的定义, 建立了概率的加法公式与乘法公式. 在结尾时又不加运算的将 5 个问题推荐给读者, 他对这些问题的解法发表于八年后的 1665 年.

所谓 "赌徒输光问题" 是最后一题.

问题 5: A 和 B 各拿 12 块筹码并用三颗骰子赌, 条件是: 如果投得 11 点, A 就给 B 一块筹码; 如果投得 14 点, B 就给 A 一块筹码. 谁先得到全部筹码的算赢. 证明 A 与 B 取胜的机会之比为 244 140 625 比 282 429 536 481.

可以看出惠更斯的确十分明确地提出了赌徒输光问题. 至于它的解答可以这样考虑: 投三颗均匀骰子得 11 点的概率为 $\dfrac{27}{216}$, 投得 14 点的概率是 $\dfrac{15}{216}$, 因此 A 胜一块筹码的概率 $p=\dfrac{15}{15+27}=\dfrac{5}{14}$, 而 B 胜一块筹码的概率为 $q=\dfrac{27}{15+27}=\dfrac{9}{14}$, 他们分别有赌本 $a=12$ 及 $b=12$. 可以通过《概率论基础》(2.3.19) 式求得 A 赢的概率 q_a, 通过 (2.3.21) 求得 B 赢的概率 p_a, 不过当时的通行术语是求 A 与 B 取胜的机会之比即 $q_a:p_a$. 当 $a=b$ 时, 该比为 $\left(\dfrac{p}{q}\right)^a$, 而问题 5 的正确答案应为 5^{12} 比 9^{12}, 这与问题 5 所求证的二数之比完全一致, 可见惠更斯已完全掌握该类问题的正确解法. 但是正如雅科布・伯努利所正确批评的, 惠更斯太关心解决数值问题而不用代数符号讨论一般形式的问题, 而后者有可能获得一般规律. 不过惠更斯的书在 19 世纪之前还一直是学习概率论的重要教材, 影响很大.

1713 年雅科布・伯努利的名著《猜度术》在他逝世 8 年之后出版. 该书共分 4 个部分. 在第一部分重印了惠更斯的上述著作, 并附上对其中大部分命题的评注. 标题为 "一篇关于惠更斯的机会

博弈可能性计算的论文, 附雅科布·伯努利的评注" 的论文, 篇幅竟达原著之五倍, 可见惠更斯对他的深刻影响以及他对惠更斯的敬重. 该书对于赌徒输光问题首次给出了一般性公式. 当然这本书最大的贡献是在最后部分证明了现在称之为伯努利大数定律的重要结果, 开创了概率论发展的一个新阶段. 伯努利的书后来也成为标准概率论教材.

二、随机游动的常返性

近代研究更关心随机游动的极限行为, 最基本的问题之一是从原点出发随机游动的质点是否一定会再回到原点来.

若以 $p_{00}^{(n)}$ 表示从原点出发的质点, 经过 n 步游动又回到原点的概率, 当

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = +\infty$$

时, 称随机游动是**常返的**.

1. 直线上无限制随机游动

在《概率论基础》第二章§3 中对直线上无限制随机游动已导出

$$p_{00}^{(n)} = \begin{cases} \dbinom{2k}{k} p^k q^k, & n = 2k \\ 0 \quad\quad\quad , & n = 2k+1 \end{cases}$$

利用斯特林公式 $k! \sim k^{k+\frac{1}{2}} \mathrm{e}^{-k} \sqrt{2\pi}$ 有

$$\frac{(2k)!}{k!k!}(pq)^k \sim \frac{(4pq)^k}{\sqrt{k\pi}}$$

由于 $pq = p(1-p) \leqslant \dfrac{1}{4}$, 等式仅当 $p = q = \dfrac{1}{2}$ 时成立. 因此当 $p \neq q$ 时,

$$\sum_{n=1}^{\infty} p_{00}^{(n)} < \infty$$

而当 $p=q$ 时,

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{k=1}^{\infty} \binom{2k}{k} \left(\frac{1}{2}\right)^{2k} \sim \sum_{k=1}^{\infty} \frac{1}{\sqrt{k\pi}} = +\infty$$

结论是直线上只有对称随机游动是常返的.

2. 平面上的随机游动

平面上的非对称随机游动更是非常返的, 因此我们只讨论对称的情况.《概率论基础》第二章 § 3 例 6 已算出

$$p_{00}^{(2n)} = \left(\frac{1}{4}\right)^{2n} \binom{2n}{n}^2$$

由斯特林公式

$$\left[\left(\frac{1}{4}\right)^{n} \binom{2n}{n}\right]^2 \sim \frac{1}{n\pi}$$

因此

$$\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{n=1}^{\infty} \left[\left(\frac{1}{4}\right)^{n} \binom{2n}{n}\right]^2 \sim \sum_{n=1}^{\infty} \frac{1}{n\pi} = +\infty$$

所以平面上对称随机游动还是常返的.

3. d 维空间格子点上的对称随机游动

这时 $p_{00}^{(2n)} \sim \dfrac{1}{n^{d/2}}$, 因此当 $d \geqslant 3$ 时,

$$\sum_{n=1}^{\infty} p_{00}^{(n)} \sim \sum_{n=1}^{\infty} \frac{1}{n^{d/2}} < \infty$$

这样一来我们惊奇地发现, 即使是对称的随机游动, 在三维及三维以上的空间, 都是非常返的. 波利亚在 1921 年首先发现这些结果, 1940 年有人算出 3 维空间中作对称随机游动的质点返回的概率仅为 0.35.

三、随机游动逼近布朗运动

随机游动描述直线上质点的运动, 每单位时间只向右或向左移动一小格.

这也可以看作是在极短间隔中对连续变化过程观察的记录. 例如每隔 0.000 1 秒观察一次某股票价格的涨落, 这样的记录, 连在一起就是一根连续的曲线. 用等间隔观察逼近连续涨落在几乎所有的学科都是有用的方法. 下面我们从随机游动模型出发, 加上若干假定, 导出一种最重要的随机过程.

以 X_i 记质点的第 i 次移动, 它们相互独立, 同服从如下分布:

$$P\{X_i=-1\}=P\{X_i=1\}=\frac{1}{2}$$

因此

$$EX_i=0,\quad DX_i=1$$

假定质点每隔 Δt 移动一次, 每次向左或向右移动一格, 长度为 Δx, 若质点于 0 时刻从 0 点出发, 则它在 t 时刻的位置为

$$W(t)=\Delta x\Big(X_1+X_2+\cdots+X_{\left[\frac{t}{\Delta t}\right]}\Big)$$

因此 $W(0)=0$, 且

$$EW(t)=0,\quad DW(t)=(\Delta x)^2\left[\frac{t}{\Delta t}\right]$$

为使讨论有意义, 对 Δt 及 Δx 的取法必须细加研究, 最后确定取

$$\Delta x=c\sqrt{\Delta t}$$

这样一来, 令 $\Delta t\longrightarrow 0$, 可得

$$EW(t)=0,\quad DW(t)\longrightarrow c^2t$$

由于 $W(t)$ 是独立同分布随机变量之和, 其方差存在, 由中心极限定理知

$$W(t)\sim N(0,c^2t)$$

这样我们知道, 在任何时刻 t, 描述质点位置的随机过程 $W(t)$ 都服从正态分布.

由随机游动的假定, 可以知道 $W(t)$ 在互不相交的时间区间的变化是相互独立的, 即 $W(t)$ 是独立增量过程, 而且在 $(t,\ t+\tau)$ 中的变化只与时间区间长度 τ 有关, 即有平稳性 (时齐性). 总而

言之, $W(t)$ 是时齐、独立增量过程, 具有正态分布, 这种随机过程称为布朗运动, 它与具有时齐 (平稳) 独立增量泊松分布的泊松过程, 是当代概率论研究中最重要的二类随机过程, 扮演着基本粒子一样的角色.

1827 年英国植物学家布朗观察到浮游在水中的花粉进行着无规则的运动. 为解释布朗运动, 许多伟大的物理学家都进行了研究, 直到 1905 年爱因斯坦才作出第一个与实验相符的说明, 可以说关于布朗运动的研究对物理学的近代发展影响巨大.

关于布朗运动数学模型的研究对概率论学科发展的影响甚至更大. 首先 1900 年在庞加莱指导下, 巴舍利耶 (Bachelier, 1870—1946) 完成了他的博士论文 "投机理论", 该文如今被誉为开创了 "连续时间随机过程" 与 "连续时间金融学" 两门学科. 在经济金融界该文因太超前而被埋没近 60 年. 在数学界该文虽未得到应有的评价, 不过, 后继的研究者还是大有人在. 1918 年维纳建立它的随机过程理论, 指出该过程样本函数以概率 1 处处连续, 处处不可微. 因此布朗运动过程在概率界有个同等流行的名字——维纳过程. 之后对该过程研究做出较大贡献的有莱维和日本的伊藤清, 特别是后者创立的 Itô 随机微积分, 现在成了连续时间金融学的重要工具, 致使最近 20 年来在概率统计界引起了一场金融热.

数学札记之十

算 术 密 度

一、算术密度

"从正整数中随机地选取两数, 求此两数互素的概率", 这是俄国数学家切比雪夫最先提出的问题. 这个问题是对古典型概率的一大挑战. 它试图保持等可能性而把概率的定义推广到自然数构成的无限样本空间 $\Omega = \{1, 2, 3, \cdots\}$.

类似的想法在几何中曾出现过, 以 1665 年牛顿的几何概率,

1763 年贝叶斯的台球桌模型和 1777 年蒲丰投针问题为代表, 后来借助于几何体的测度 (长度、面积、体积) 建立了几何概率, 取得了一定的成功.

现在的问题是从数论中关于既约分数而提出的.

对于无限集合 Ω, 要保证随机选取的等可能性, 取个别数的概率只能为 0, 取预先指定的任何有限个数的概率也应为 0. 取得偶数呢? 似乎应该是 $\frac{1}{2}$; 取被 3 整除的数呢? 似乎应该是 $\frac{1}{3}$; $\cdots\cdots$. 一般地, 取到能被 k 整除的数 (其全体即为 k 的倍数 E_k) 的概率应为 $\frac{1}{k}$.

为了使上述想法合理化, 可以考虑有限样本空间 $\Omega_N = \{1, 2, \cdots, N\}$, 对自然数集 A, 定义

$$D_N(A) = \frac{|A \cap \Omega_N|}{N} \tag{1}$$

其中 $|B|$ 表示集合 B 的计数.

定义 若 $\lim\limits_{N\to\infty} D_N(A)$ 存在, 记为 $P(A)$, 称为数集 A 的**算术密度**.

按此定义

$$P(E_k) = \lim_{N\to\infty} \frac{|E_k \cap \Omega_N|}{N} = \frac{1}{k} \tag{2}$$

与直观相符, 而且取到个别数或预先指定的有限个数的概率都为 0. 进一步由数论中的欧拉定理知道任取一数为素数的概率也为 0.

仿筛法可知

$$P(E_2 \cup E_3) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = P(E_2) + P(E_3) - P(E_6)$$

一般地, 对 E_{k_1}, E_{k_2} 成立加法公式

$$P(E_{k_1} \cup E_{k_2}) = P(E_{k_1}) + P(E_{k_2}) - P(E_{k_1\times k_2}) \tag{3}$$

可以发现素数有特殊性, 若 p_1 与 p_2 都是素数, 则

$$P(E_{p_1\times p_2}) = P(E_{p_1})P(E_{p_2})$$

这时 $E_{p_1 \times p_2} = E_{p_1} \cap E_{p_2}$, 故成立

$$P(E_{p_1} \cap E_{p_2}) = P(E_{p_1}) \times P(E_{p_2}) \tag{4}$$

因此 E_{p_1} 与 E_{p_2} "独立".

这也可以推广到 n 个素数 $p_1, p_2, \cdots, p_n$ 的场合.

一般地, 若 $k_1, k_2, \cdots, k_m$ 互素, 也有

$$P(E_{k_1} \cap E_{k_2} \cap \cdots \cap E_{k_m}) = P(E_{k_1})P(E_{k_2}) \cdots P(E_{k_m}) \tag{5}$$

二、可列可加性

在概率论的公理化结构中, 如 Ω 是离散样本空间, $\mathscr{F}$ 可取为 Ω 的子集全体. 在算术密度中却遇到了麻烦, 因为上面已得到, 对任何一个正整数 k, 作为单点集 $\{k\}$, 应有

$$P\big(\{k\}\big) = 0$$

若概率 P 具有可列可加性, 那么

$$P(\Omega) = \sum_{k=1}^{\infty} P\big(\{k\}\big) = 0$$

这就与概率的规范性矛盾, 而且这时任何集合的概率皆为 0, 还讨论什么?

因此在算术密度研究中, 概率空间 $(\Omega, \mathscr{F}, P)$ 的给定, 要么放弃 P 的可列可加性, 只保留有限可加性; 要么保留 P 的可列可加性, 把 $\mathscr{F}$ 加以限定.

三、有限场合

数论中最基本的是可除性理论, 这时 $\mathscr{F}$ 可以设为由有限个 $E_{k_1}, E_{k_2}, \cdots, E_{k_n}$ 产生的 σ 域.

问题 随机选取一正整数, 求它不被 5 整除, 只能被 3 整除同时又能被 4 或 6 整除的概率.

这是见诸于钟开莱《初等概率论附随机过程》书中的一道例题, 曾被国内若干教科书所沿用.

按题意, 要求 $E_3\overline{E}_5(E_4\cup E_6)$, 容易证明它等价于 $E_2E_3\overline{E}_5$, 因此

$$\begin{aligned}P\{E_3\overline{E}_5(E_4\cup E_6)\}&=P\{E_2E_3\overline{E}_5\}\\&=P\{E_2E_3\}-P\{E_2E_3E_5\}\\&=\frac{1}{6}-\frac{1}{30}=\frac{2}{15}\end{aligned}\tag{6}$$

这时 $\mathscr{F}$ 只由 E_2,E_3,E_4,E_5,E_6 产生, 只用到概率的有限可加性, 整个解题过程是严格的.

四、切比雪夫问题

回到本文开头提到的切比雪夫问题上来, 这就是《概率论基础》的习题二 **50 题, 不过出示了答案 $\frac{6}{\pi^2}$, 改成证明题.

其解题思路是: 若以 A_p 记取出的第一个数能被 p 整除, 则 $P(A_p)=\frac{1}{p}$, 以 B_p 记取出的第二个数能被 p 整除, 则 $P(B_p)=\frac{1}{p}$, 以 C 记两数互素, 则

$$C=\bigcap_{p素数}\left(\overline{A_pB_p}\right)\tag{7}$$

即所取两数不能同时被某个素数整除.

利用前述素数的特殊性 (5) 及 A_p与 B_p 的独立性

$$\begin{aligned}P(C)&=\prod_{p素数}P(\overline{A_pB_p})\\&=\prod_{p素数}\left[1-P(A_pB_p)\right]\\&=\prod_{p素数}\left[1-P(A_p)P(B_p)\right]\\&=\prod_{p素数}\left(1-\frac{1}{p^2}\right)\\&=\frac{1}{\zeta(2)}=\frac{6}{\pi^2}\end{aligned}\tag{8}$$

详细的解法参看习题解答部分. 这种问题通常都会碰上黎曼 ζ 函数 $\zeta(s)$, 这里 $s=2$, 用 $\zeta(2)=\dfrac{\pi^2}{6}$ 得到答案.

这个解答过程中实际上利用了概率的可列可加性, 因此不能算作完全严格. 不过这个问题可用数论方法严格处理. 参看维诺格拉陀夫《数论基础》第二章问题 21.

五、欧拉函数的统计学

欧拉函数 $\varphi(n)$ 是数论中重要的函数, 有众多应用, 本身也很有趣.

定义 欧拉函数 $\varphi(n)$ 定义为不超过 n 且与 n 互素的正整数的个数.

例如

$$\varphi(1)=1,\quad \varphi(2)=1,\quad \varphi(3)=2,\quad \varphi(4)=2,\quad \varphi(5)=4,$$
$$\varphi(6)=2,\quad \varphi(7)=6,\quad \varphi(8)=4,\quad \varphi(9)=6,\quad \varphi(10)=4$$

一般地, 对素数 p,

$$\varphi(p)=p-1,\quad \varphi(p^\alpha)=p^\alpha-p^{\alpha-1} \tag{9}$$

若 n 的标准分解式为

$$n=p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_k^{\alpha_k}$$

则

$$\varphi(n)=n\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_k}\right) \tag{10}$$

欧拉函数 $\varphi(n)$ 有一个重要的性质.

性质 (欧拉函数的可乘性): 若 m 与 n 互素, 则

$$\varphi(mn)=\varphi(m)\varphi(n) \tag{11}$$

这性质显然可推广到多个互素数的场合.

若引入函数 $\rho_p(n)$ 如下:

$$\rho_p(n)=\begin{cases}1, & p\mid n\\ 0, & p\nmid n\end{cases} \tag{12}$$

这里 $p\mid n$ 表示 p 可整除 n, 而 $p\nmid n$ 则表示 p 不能整除 n.

若以 ε_i 记 0 或 1, 则

$$\begin{aligned}&P\{\rho_{p_1}(n)=\varepsilon_1,\cdots,\rho_{p_k}(n)=\varepsilon_k\}\\ &\quad=P\{\rho_{p_1}(n)=\varepsilon_1\}\cdots P\{\rho_{p_k}(n)=\varepsilon_k\}\end{aligned} \tag{13}$$

这是以前讲过的, 任选一数 n 能否被素数 $p_1,\cdots,p_k$ 整除是相互独立的, 即 (5) 的特例. 因此 (13) 式表明 $\rho_{p_1}(n),\cdots,\rho_{p_k}(n)$ 是相互独立的.

利用新记号

$$\frac{\varphi(n)}{n}=\prod_p\left(1-\frac{\rho_p(n)}{p}\right) \tag{14}$$

这里 p 跑遍一切小于 n 的素数.

对于任意定义在正整数上的函数 $f(n)$, 定义 $f(n)$ 的**均值**

$$M\{f(n)\}=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}f(n) \tag{15}$$

只要极限存在.

算术密度的定义可以看作它的特例.

由 (13) 式可得

$$\begin{aligned}M\Big\{\prod_{p\leqslant p_k}\left(1-\frac{\rho_p(n)}{p}\right)\Big\}&=\prod_{p\leqslant p_k}M\Big\{\left(1-\frac{\rho_p(n)}{p}\right)\Big\}\\ &=\prod_{p\leqslant p_k}\left(1-\frac{1}{p^2}\right)\end{aligned}$$

它提示

$$
\begin{aligned}
M\left\{\frac{\varphi(n)}{n}\right\} &= M\left\{\prod_p\left(1-\frac{\rho_p(n)}{p}\right)\right\} \\
&= \prod_p M\left\{\left(1-\frac{\rho_p(n)}{p}\right)\right\} \\
&= \prod_p\left(1-\frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}
\end{aligned}
$$

这里还是遇到可列可加性的困难, 不算严格推导. 不过, 它也可以严格处理, 而且还有更深入的研究, 见 Kac 书.

第三章　随机变量与分布函数

章前引言

从本章起变换语言,用随机变量描述事件,以分布函数给出概率,从而使概率论与其他数学学科对接,内容的展开明显加快.主角换成分布函数.

在上章 §4 的应用实例中出现的 μ 就是随机变量,可见随机变量概念是十分自然的.事实上,在 19 世纪中叶之后随机变量概念出现,随后即被广泛采用就使概率论进入现代形式.

陆续介绍了一些分布函数,有些是老相识,有些是新面孔.最重要的分布——正态分布终于登场,此后一直扮演主角,而均匀分布则为几何概率提供精确的数学语言.

从一维到多维,从变量到函数,研究内容大为丰富.

课文导读

§3.1　随机变量及其分布

用随机变量来描述随机现象是当代概率统计的通用做法,概念直观,方法自然,与数学其他分支容易建立联系,有助于概率论成为数学大家庭中的一员.

从此之后,事件一般都通过随机变量的某一个关系式表出,而概率则用分布函数计算.

在公理化体系中, 随机变量 $\xi(\omega)$ 定义为样本空间 Ω 到欧几里得空间 $\mathbf{R}^1$ 的可测映照, 其随机性由 ω 表出, 在试验前不知哪一个 ω 将出现, "变量" 二字表明试验的结果用一个实数表示, 把概率论问题 "数值化", 而可测性 $\{\omega:\xi(\omega)\in B\}\in\mathscr{F}$ 则为分布函数 $F(x)=P\{\xi<x\}$ 的定义铺平道路. 在习题 *49, *50 中证明可测性也能用表面更宽的条件 $\{\omega:\xi(\omega)<x\}\in\mathscr{F}$ 代替, 这实质上已是测度论内容, 并不要求本书读者必须掌握, 但有兴趣的读者不妨阅读教学札记之十九 "概率论公理化结构与测度论", 增加了解.

随机变量依取值的不同分成离散型与连续型, 前者用概率分布或分布列描述, 后者用密度函数描述, 分布函数统一了二者.

分布函数是非降、取值 $[0,1]$ 的左连续函数, 它完整描述了随机变量, 又便于处理, 因此成为今后研究的主要对象.

概率论中最有名的分布集中在本节出现. 其中离散型的已在上面两章亮过相, 而连续型的则属首次登场, 尤以正态分布和均匀分布最为重要.

《概率论基础》特意安排让离散型等待时间分布: 几何分布, 帕斯卡分布, 负二项分布与连续型等待时间分布: 指数分布, 埃尔朗分布, Γ 分布对称出场. 前者在第二章通过伯努利试验引入, 后者则经由泊松过程在本章导出. 这两列分布性质上的相似性实非偶然, 因为泊松过程可由伯努利试验逼近. 与此有关的内容以后多次在正文与习题出现, 也可参看教学札记之十二 "关于两个等待时间分布序列".

正态分布是概率论中最重要的分布, 在全书的黄金分割时段首次出现, 以后章节还将对它作深入讨论. 读者在这个阶段可主要练习它的有关计算. 若手边没有正态分布表, Shah (1985) 推荐的如下近似公式可达很高精度:

$$\Phi(z)=\frac{z(4.4-z)}{10}+\frac{1}{2}+\varepsilon,\quad 0\leqslant z\leqslant 2.2,\quad |\varepsilon|<0.005$$

示性函数连接了事件与随机变量, 使上两章的许多结果可转

换成随机变量的结论.

勒贝格分解中的奇异型分布最重要的例子来自康托尔奇异函数, 参阅教学札记之二十七 "[0,1] 中数的 g 进制展开与概率论".

§ 3.2 随机向量, 随机变量的独立性

随机向量即多维随机变量, 对它的研究不但涉及每个分量自身 (边际分布), 更要考察它们之间的全部关系 (联合分布函数), 因此条件分布、独立性 (几乎完全平行于事件场合) 应需而生, 大大丰富了研究的内容.

联合分布函数以及它的两种特殊表现形式 —— 多元概率分布与多元密度函数成为主要研究工具.

有名的一元分布众多, 但有名的多元分布则寥寥无几.

在离散型场合, 主要只有两种, 它们是作为二项分布与超几何分布推广的多项分布与多元超几何分布. 当摸球模型袋中的球从二色改成多色时作此推广. 多项分布对应放回摸球, 这种摸球方式中每次摸球结果独立同分布; 多元超几何分布对应不放回摸球, 这种摸球方式中每次摸球结果同分布 (回忆抽签与顺序无关!) 但不独立. 认识放回与不放回摸球的差别在概率论中十分重要, 因此在随机向量场合重又提起,《概率论基础》表 3.2.1 及表 3.2.2 给出简单实例, 教学札记之十三 "摸球与抽样" 中作进一步概括. 通过合并不同颜色球为一类 (降低了问题的维数), 可得到边际分布的有关结果, 这是处理这两种分布的重要技巧, 参看习题三中 23 及 24 题.

二元均匀分布可概括几何概率的大部分问题, 到了这里总算可以把几何概率一节的各个例子的潜在假设完全弄清.

二元正态扮演本节主角. 它是应用最多的二元分布, 且具有许多特殊的性质. 本节引进的典型分解法使讨论大为简便. 特别值得指出的是条件分布仍为正态, 且其均值 $\mu=\mu_2+\rho\dfrac{\sigma_2}{\sigma_1}(x-\mu_1)$ 是 x 的线性函数, 方差 $\sigma_2^2(1-\rho^2)$ 是常数, 与 x 无关, 这个性质在应用

中十分重要, 图 3-1 有助于理解这一结论.

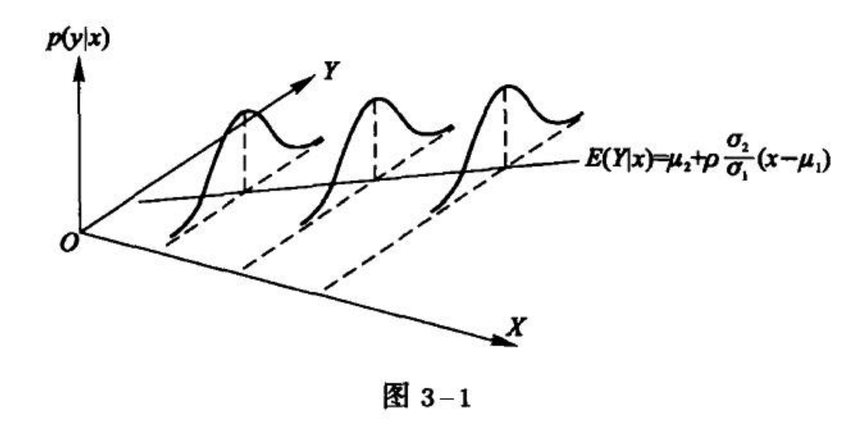


图 3-1

高斯用最大似然法导出正态分布是对概率论的重大贡献.《概率论基础》的推导因利用柯西方程, 比原法略微初等.

关于正态分布的其他刻画可参看教学札记之十四 "正态分布的两种刻画".

§ 3.3 随机变量的函数及其分布

本节讲清了在什么情况下随机变量的函数还是随机变量, 并给出了计算它们的分布的各种方法. 这是概率论通向应用 (特别是数理统计) 的重要路口, 也是对学生们进行训练的重要操场.

总的说, 求分布问题, 对离散型易, 对连续型难. 因此本节由连续型唱主角.

本节布局以方法为主, 以分布为辅. 求随机变量函数分布的方法是本节的主题, 当然应占主位. 但是在介绍这些方法时一定要举许多例子, 这类例子俯拾即是, 因此就给教科书编写者很大的选择余地. 《概率论基础》利用这个机会选择典型题型, 介绍各种分布, 特别注意引入新分布以及揭示分布之间的联系, 大致做到每道例题都有所指. 请读者在全节读完后复核.

题材的展开以三个类型 (一对一, 多对一, 多对多) 对两种方法 (直接法, 变换法) 交叉的方式进行, 以公式、例题或专题的形式给出, 概要列于下表:

	直接法	变换法
$\eta = g(\xi)$	例 1—例 4 均匀分布特殊性	例 5
$\eta = g(\xi_1, \xi_2)$	和的公式与商的公式 极小值—极大值	增补变量法: 例 8
$\eta_1 = g_1(\xi_1, \xi_2)$ $\eta_2 = g_2(\xi_1, \xi_2)$		例 6—例 10

在一对一场合, 两种方法几乎等价, 在变换法适用的单调或分段单调的情形, 用直接法还更具直观, 因此这种场合还是直接法用得多.

在多对一场合, 初看使用变换法是不可能的, 因此只能用直接法. 但是可以采用增补变量化为多对多, 还是可以使用变换法. 而且由于增补的变量可以按需要选择 (见例 8) 或选择最简单的变量, 所以进行起来还颇为方便. 事实上, 直接法的和、差、积、商的公式都可以用增补变量法方便导出, 读者不妨以此作为练习题, 以加深理解.

在多对多场合, 直接法不易确定积分区域, 多重积分计算也过于艰难, 这时多数采用变换法.

学习本节, 单看不行, 要亲自动手, 这也是考验微积分水平的大好场所. 习题三 25 题之后的习题大都是精选的求随机变量函数分布的典型题. 教学札记之二十六 "统计学三大分布的推导" 介绍了统计学三大分布的各种推导法, 对理解本节内容或有助益.

通过本节课文及习题至少引入对数正态, 柯西, 瑞利, χ^2, t, F, 拉普拉斯, 韦布尔, 麦克斯韦, β 等十个新分布, 也提供大量关于分布之间联系的信息, 请同学们自己总结一下.

习题解答与评注

1. 直线上有一质点, 每经一个单位时间, 它分别以概率 p 及 q

向右或向左移动一格, 若该质点在时刻 0 从原点出发, 而且每次移动是相互独立的, 试用随机变量来描述这质点的运动 (以 S_n 表示时刻 n 时质点的位置).

解 以 S_n 表示时刻 n 时质点的位置, 则若在时刻 n 该质点向右移动了 k 次, 必向左移动 $n-k$ 次, 因此 S_n 的概率分布为

$$P\{S_n = 2k-n\} = \begin{cases} \binom{n}{k} p^k q^{n-k}, & k=0,1,2,\cdots,n \\ 0 & , \text{其他} \end{cases}$$

【评注】对比第二章 §3 关于无限制随机游动的讨论.

2. 设 ξ 为伯努利试验中第一个**游程** (连续的成功或失败) 的长, 试求 ξ 的概率分布.

提示: 几何分布描述了连续的失败次数.

答 $P\{\xi = k\} = p^k q + q^k p,\ k = 1,2,3,\cdots$.

【评注】在统计中, 游程用于检验某类现象的随机性.

3. C 应取何值才能使下列数列成为概率分布:

(1) $p_k = \dfrac{C}{N}, \quad k = 1,2,\cdots,N;$

(2) $p_k = C\dfrac{\lambda^k}{k!}, \quad k=1,2,\cdots, \quad \lambda > 0.$

提示: 记住概率分布的充要条件.

答 (1) $C=1$; (2) $C = \left(\mathrm{e}^{\lambda}-1\right)^{-1}$.

【评注】(1) 这个分布常称为离散均匀分布.

(2) 若 $\xi \sim P(\lambda)$, 则 $p_k = P\{\xi = k \mid \xi > 0\},\ k=1,2,\cdots$.

本题提醒大家, 对离散型概率分布 p_k, 应注意 k 的取值范围, 最好养成良好习惯, 在书写概率分布时, 同时写明取值范围.

4. 若分布函数定义为 $F(x) = P\{\xi \leqslant x\}$, 试证这时的 $F(x)$ 具有下列性质:

(i) 非降; (ii) $F(-\infty)=0$, $F(+\infty)=1$; (iii) 右连续.

提示: 逐句逐字模仿课文进行证明, 但要特别注意在哪些地方必须做出更改.

【评注】分布函数的另一种常见定义. 阅读概率论书刊时必须注意.

本题设置目的有二. 第一, 指出有另一种稍微不同的定义存在, 使用或阅读时当加以注意; 第二, 抄一遍证明, 明白为什么有左连续与右连续的区别, 以后计算离散型概率时自会当心.

5. 若 $\zeta \sim N(0,1)$, 试求常数 a, b, c 使 (1) $a = P\{\zeta \geqslant 1.645\}$; (2) $P\{|\zeta| < b\} = 95\%$; (3) $P\{|\zeta - c| > c\} = 0.51$.

提示: 直接利用《概率论基础》附录三标准正态分布函数的数值表.

答 (1) 0.05; (2) 1.96; (3) 1.165.

【评注】正态分布计算训练之一: 查表.

6. 妊娠天数 ξ 的分布函数为 $N(270, 100)$, 求 ξ 落在下列范围的概率: (1) (260,280); (2) 短于 250 天; (3) 长于 300 天.

提示: 先化为标准正态分布后查《概率论基础》附录三.

答 (1) 0.682 69; (2) 0.022 75; (3) 0.001 35.

【评注】正态分布计算训练之二: 化为标准正态分布.

7. 若 ξ 的分布函数为 $N(60, 9)$, 求分点 x_1, x_2, x_3, x_4, 使 ξ 落在 $(-\infty, x_1)$, (x_1, x_2), (x_2, x_3), (x_3, x_4), $(x_4, +\infty)$ 中的概率之比为 $7:24:38:24:7$.

提示: 对标准正态分布找满足所给概率比的分点 z_1, z_2, z_3, z_4, 再转换成 x_1, x_2, x_3, x_4.

答 $x_1 = 55.5$, $x_2 = 58.5$, $x_3 = 61.5$, $x_4 = 64.5$.

【评注】正态分布计算训练之三: 化标准, 倒查表.

*8. 在帕斯卡分布 $\binom{k-1}{r-1} p^r q^{k-r}$ 中, 令 $k\Delta t = t$, $p = \lambda \Delta t$, 试证当 $\Delta t \to 0$ 时, 它能用 $\dfrac{\lambda(\lambda t)^{r-1}}{(r-1)!} \mathrm{e}^{-\lambda t} \cdot \Delta t$ 来逼近. (这可以解释为第 r 次成功发生在 $(t,\ t + \Delta t)$ 中的概率, 其密度函数正好是参数为 r 的埃尔朗分布. 用这种方法可以把课文中的对比严格化.)

证明 帕斯卡分布

$$
\begin{aligned}
&\binom{k-1}{r-1}p^r q^{k-r}\\
&=\frac{(k-1)(k-2)\cdots(k-r+1)}{(r-1)!}\lambda^r(\Delta t)^r\big(1-\lambda\Delta t\big)^{k-r}\\
&=\Delta t\cdot\frac{(k\Delta t-\Delta t)\cdots\big(k\Delta t-(r-1)\Delta t\big)\lambda^r}{(r-1)!\big(1-\lambda\Delta t\big)^r}\big(1-\lambda\Delta t\big)^k\\
&=\Delta t\cdot\frac{(t-\Delta t)\cdots\big(t-(r-1)\Delta t\big)\lambda^r}{(r-1)!\big(1-\lambda\Delta t\big)^r}\big(1-\lambda\Delta t\big)^{\frac{t}{\Delta t}}
\end{aligned}
$$

而

$$
\lim_{\Delta t\to 0}\big(1-\lambda\Delta t\big)^{\frac{t}{\Delta t}}=\mathrm{e}^{-\lambda t}
$$

因此对固定 r,

$$
\begin{aligned}
&\lim_{\Delta t\to 0}\frac{(t-\Delta t)\cdots\big(t-(r-1)\Delta t\big)\lambda^r}{(r-1)!\big(1-\lambda\Delta t\big)^r}\big(1-\lambda\Delta t\big)^{\frac{t}{\Delta t}}\\
&\qquad=\frac{\lambda\big(\lambda t\big)^{r-1}}{(r-1)!}\mathrm{e}^{-\lambda t}
\end{aligned}
$$

所以

$$
\binom{k-1}{r-1}p^r q^{k-r}=\frac{\lambda\big(\lambda t\big)^{r-1}}{(r-1)!}\mathrm{e}^{-\lambda t}\cdot\Delta t+o(\Delta t)
$$

于是, 当 $\Delta t \to 0$ 时, 帕斯卡分布 $\binom{k-1}{r-1}p^r q^{k-r}$ 能用 $\frac{\lambda(\lambda t)^{r-1}}{(r-1)!}\mathrm{e}^{-\lambda t}\cdot\Delta t$ 来逼近.

【评注】第五章 §2 还会回到这个问题.

9. 在生存分析中, 作为研究对象的是非负随机变量, 它们的分布称为寿命分布. 若 ξ 是非负随机变量, 其分布函数为 $F(x)$, 密度函数为 $f(x)$, 这时通常还引入**生存函数** $S(x)=P\{\xi\geqslant x\}$ 及**失效率函数** $\lambda(x)=\dfrac{f(x)}{1-F(x)}$. 试导出 $S(x)$, $\lambda(x)$, $F(x)$ 及 $f(x)$ 之间的

关系式, 并以指数分布验证之.

解

$$S(x) = 1 - F(x)$$

$$S(x) = \int_x^{\infty} f(t)\,\mathrm{d}t$$

$$S'(x) = -f(x)$$

$$\lambda(x) = \frac{f(x)}{1 - F(x)} = -\frac{S'(x)}{S(x)} \text{ 或 } \frac{S'(x)}{S(x)} = -\lambda(x)$$

$$\int_0^x \frac{S'(t)}{S(t)}\,\mathrm{d}t = -\int_0^x \lambda(t)\,\mathrm{d}t$$

$$\ln S(x) = -\int_0^x \lambda(t)\,\mathrm{d}t$$

$$\therefore\ S(x) = \mathrm{e}^{-\int_0^x \lambda(t)\,\mathrm{d}t}$$

$$F(x) = 1 - S(x) = 1 - \mathrm{e}^{-\int_0^x \lambda(t)\,\mathrm{d}t}$$

$$f(x) = F'(x) = \lambda(x)\exp\left[-\int_0^x \lambda(t)\,\mathrm{d}t\right]$$

对指数分布 Exp (λ),

$$f(x) = \lambda\,\mathrm{e}^{-\lambda x}$$

$$F(x) = 1 - \mathrm{e}^{-\lambda x}$$

$$S(x) = \mathrm{e}^{-\lambda x}$$

$$\lambda(x) = \lambda$$

【评注】在生存分析, 可靠性等学科中, 这几个是标准术语.

顺便指出, 失效率函数为常数是指数分布的刻画性性质, 有人据此对指数分布的常见性提出“浴盆模型”作为解释. 生物体的死亡或元器件的失效大都发生在早期或晚期. 有先天性缺陷者早期就出问题, 之后进入一个平稳期, 到了晚期因老化又纷纷出毛病, 所以失效率函数先有一段快速下降, 再有一长段持平, 最后又快速上升, 形如浴盆, 论者认为, 指数分布描述了中间持平的一段. 电话通

话时间也如是，因此排队论中假设服务时间为指数分布导出的公式在实际应用中效果良好.

10. 设随机变量 ξ 取值于 $[0,1]$，若 $P\{x \leqslant \xi < y\}$ 只与长度 $y-x$ 有关 (对一切 $0 \leqslant x \leqslant y \leqslant 1$)，试证 ξ 服从 $[0,1]$ 均匀分布.

证 记 $P\{x \leqslant \xi < y\} = f(y-x)$，则对 $x=0, \forall y \in [0,1]$，有

$$P\{0 \leqslant \xi < y\} = f(y)$$

对 $\forall y_1, y_2 \in [0,1], y_1 < y_2$,

$$P\{0 \leqslant \xi < y_1 + y_2\} = P\{0 \leqslant \xi < y_1\} + P\{y_1 \leqslant \xi < y_1 + y_2\}$$

即有

$$f(y_1 + y_2) = f(y_1) + f(y_2)$$

故

$$f(y) = Cy$$

由 $f(1) = f(1-0) = P\{0 \leqslant \xi < 1\} = 1$ 推得 $C=1$，所以 $f(x)=x$，即

$$P\{0 \leqslant \xi < x\} = x, \quad x \in [0,1]$$

因而 ξ 服从 $[0,1]$ 均匀分布.

【评注】联系几何概率，刻画均匀分布.

*11. 若存在 Θ 上的实值函数 $Q(\theta)$ 及 $D(\theta)$ 以及 $T(x)$ 及 $S(x)$，使

$$f_\theta(x) = \exp\{Q(\theta)T(x) + D(\theta) + S(x)\}$$

则称 $\{f_\theta,\ \theta \in \Theta\}$ 是一个**单参数的指数族**. 证明 (1) 正态分布 $N(m_0, \sigma^2)$，已知 m_0，关于参数 σ; (2) 正态分布 $N(m, \sigma_0^2)$，已知 σ_0，关于参数 m; (3) 泊松分布 $P(\lambda)$ 关于 λ 是一个单参数的指数族.

但是 $[0,\theta]$ 上均匀分布，关于 θ 不是一个单参数的指数族.

证 (1) 该正态分布密度函数

$$
\begin{aligned}
f_\sigma(x) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-m_0)^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}(x-m_0)^2 + \ln\frac{1}{\sqrt{2\pi}\sigma}\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}(x-m_0)^2 - \ln\sigma - \ln\sqrt{2\pi}\right\}
\end{aligned}
$$

令

$Q(\sigma) = -\dfrac{1}{2\sigma^2}$, $T(x) = (x-m_0)^2$, $D(\sigma) = -\ln\sigma$, $S(x) = -\ln\sqrt{2\pi}$,

则有 $f_\sigma(x) = \exp\left\{Q(\sigma)T(x) + D(\sigma) + S(x)\right\}$. 所以正态分布 $N(m_0, \sigma^2)$ 是单参数 σ 的指数族.

(2), (3) 留给读者自证.

指数族 $f_\theta(x) > 0$, 均匀分布不满足.

【评注】在数理统计中, 一些重要理论结果是对指数族分布建立的. 指数族分布范围甚广, 你不妨自己再找几个.

12. 定义二元函数

$$
F(x,y) = \begin{cases} 1, & x+y > 0 \\ 0, & x+y \leqslant 0 \end{cases}
$$

验证此函数对每个变元非降, 左连续, 且满足分布函数性质 (ii), 但无法使 (3.2.5) 保持非负.

解 易验证对每个变元非降, 左连续, 且满足分布函数性质 (ii).

但若取 $a_1 = a_2 = 0$ 和 $b_1 = b_2 = 1$, 则有

$$
F(1,1) - F(1,0) - F(0,1) + F(0,0) = 1 - 1 - 1 + 0 = -1
$$

【评注】简单而重要的反例, 配合正文.

13. 若 $f_1(x)$, $f_2(y)$ 为分布密度, 为使 $f(x,y) = f_1(x)f_2(y) + h(x,y)$ 成为密度函数, $h(x,y)$ 必须而且只须满足什么条件?

提示: 利用分布密度函数的充要条件.

答 $h(x,y) \geqslant -f_1(x)f_2(y)$ 且 $\iint h(x,y)\,\mathrm{d}x\mathrm{d}y = 0$.

【评注】 帮助记忆二元密度函数的充要条件.

14. 若 $f_1(x)$, $f_2(x)$, $f_3(x)$ 是对应于分布函数 $F_1(x)$, $F_2(x)$, $F_3(x)$ 的密度函数, 证明对于一切 $\alpha(-1<\alpha<1)$, 下列函数是密度函数, 且具有相同的边际密度函数 $f_1(x)$, $f_2(x)$, $f_3(x)$:

$$
\begin{aligned}
&f_\alpha(x_1,x_2,x_3)\\
&\quad = f_1(x_1)f_2(x_2)f_3(x_3)\{1+\alpha[2F_1(x_1)-1][2F_2(x_2)-1][2F_3(x_3)-1]\}
\end{aligned}
$$

解 首先,

$$0 \leqslant F_i(x_i) \leqslant 1$$

$$-1 \leqslant 2F_i(x_i) - 1 \leqslant 1$$

$$-1 \leqslant [2F_1(x_1)-1][2F_2(x_2)-1][2F_3(x_3)-1] \leqslant 1$$

代入 $f_\alpha(x_1,x_2,x_3)$ 的表达式得

$$f_\alpha(x_1,x_2,x_3) \geqslant 0 \tag{1}$$

其次

$$
\begin{aligned}
\int_{-\infty}^{\infty}[2F_i(x_i)-1]f_i(x_i)\mathrm{d}x_i &= \int_{-\infty}^{\infty}[2F_i(x_i)-1]\mathrm{d}F_i(x_i)\\
&= \left[F_i^2(x_i)-F_i(x_i)\right]\Big|_{-\infty}^{\infty} = 0
\end{aligned}
$$

所以

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\alpha(x_1,x_2,x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\\
&= \int_{-\infty}^{\infty} f_1(x_1)\mathrm{d}x_1\int_{-\infty}^{\infty} f_2(x_2)\mathrm{d}x_2\int_{-\infty}^{\infty} f_3(x_3)\mathrm{d}x_3 = 1
\end{aligned}
\tag{2}
$$

由 (1), (2) 知 $f_\alpha(x_1,x_2,x_3)$ 是密度函数. 用类似的方法计算可得边际密度函数为

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\alpha(x_1,x_2,x_3)\mathrm{d}x_2\mathrm{d}x_3 = f_1(x_1)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\alpha(x_1,x_2,x_3)\mathrm{d}x_1\mathrm{d}x_3 = f_2(x_2)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_\alpha(x_1,x_2,x_3)\mathrm{d}x_1\mathrm{d}x_2 = f_3(x_3)$$

【评注】以一维密度函数为边际密度构造多元密度函数的方法之一.

15. 若 (ξ,η) 的联合概率分布为

η \ ξ	-1	0	1
-1	a	0	0.2
0	0.1	b	0.1
1	0	0.2	c

且 $P\{\xi\eta\neq 0\}=0.4$, $P\{\eta\leqslant 0\,|\,\xi\leqslant 0\}=\dfrac{2}{3}$, 试求:

(1) a, b, c 之值;

(2) ξ 及 η 的边际概率分布;

(3) $\xi+\eta$ 的概率分布.

解 (1) 由联合分布性质 $a+0.2+0.1+b+0.1+0.2+c=1$ 得 $a+b+c=0.4$.

由 $P\{\xi\eta\neq 0\}=a+0.2+c=0.4$ 得 $b=0.2$.

由 $P\{\eta\leqslant 0\,|\,\xi\leqslant 0\}=\dfrac{P\{\xi\leqslant 0,\eta\leqslant 0\}}{P\{\xi\leqslant 0\}}=\dfrac{a+b+0.1}{a+b+0.3}=\dfrac{2}{3}$, 得 $a+b=0.3$, 推得 $a=0.1$, 进而有 $c=0.1$.

(2)

ξ	-1	0	1
P	0.2	0.4	0.4

η	-1	0	1
P	0.3	0.4	0.3

(3)

$\xi+\eta$	-2	-1	0	1	2
P	0.1	0.1	0.4	0.3	0.1

【评注】以一题而作列联表全般训练.

16. 若 (ξ,η) 的密度函数为

$$p(x,y)=\begin{cases}A\,\mathrm{e}^{-(2x+y)}, & x>0,y>0\\ 0, & \text{其他}\end{cases}$$

试求: (1) 常数 A; (2) $P\{\xi<2,\eta<1\}$; (3) ξ 的边际分布函数; (4) $P\{\xi+\eta<2\}$; (5) $p(x\,|\,y)$; (6) $P\{\xi<2\,|\,\eta<1\}$.

提示: 独立, 双指数分布, 二元场合, 用密度函数计算种种概率.

答 (1) $A=2$; (2) $\left(1-\mathrm{e}^{-4}\right)\left(1-\mathrm{e}^{-1}\right)$; (3) $2\,\mathrm{e}^{-2x},\ x>0$; (4) $\left(1-\mathrm{e}^{-2}\right)^2$; (5) $2\,\mathrm{e}^{-2x},\ x>0$; (6) $1-\mathrm{e}^{-4}$.

【评注】以一题而作连续型的全部计算.

17. 若 $P\{\mu=m,\nu=n\}=\dfrac{(\lambda p)^m(\lambda-\lambda p)^{n-m}}{m!(n-m)!}\,\mathrm{e}^{-\lambda}$,

$m=0,1,2,\cdots,n,\ n=0,1,2,\cdots$, 试求:

(1) $P\{\nu=n\}$; (2) $P\{\mu=m\}$;

(3) $P\{\mu=m\,|\,\nu=n\}$; (4) $P\{\nu-\mu=k\}$.

解 (1)

$$\begin{aligned}P\{\nu=n\}&=\sum_{m=0}^{n}\frac{(\lambda p)^m(\lambda-\lambda p)^{n-m}}{m!(n-m)!}\,\mathrm{e}^{-\lambda}\\&=\sum_{m=0}^{n}\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda}\cdot\frac{n!}{m!(n-m)!}\,p^m(1-p)^{n-m}\\&=\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda},\qquad n=0,1,2,\cdots\end{aligned}$$

(2)

$$\begin{aligned}P\{\mu=m\}&=\sum_{n=m}^{\infty}\frac{(\lambda p)^m(\lambda-\lambda p)^{n-m}}{m!(n-m)!}\,\mathrm{e}^{-\lambda}\\&=\sum_{k=0}^{\infty}\frac{(\lambda p)^m(\lambda-\lambda p)^k}{m!k!}\,\mathrm{e}^{-\lambda}\\&=\sum_{k=0}^{\infty}\frac{(\lambda p)^m}{m!}\,\mathrm{e}^{-\lambda p}\,\frac{(\lambda-\lambda p)^k}{k!}\,\mathrm{e}^{-(\lambda-\lambda p)}\\&=\frac{(\lambda p)^m}{m!}\,\mathrm{e}^{-\lambda p},\qquad m=0,1,2,\cdots\end{aligned}$$

(3) $P\{\mu=m|\nu=n\}=\dfrac{P\{\mu=m,\nu=n\}}{p\{\nu=n\}}$

$$=\frac{(\lambda p)^m(\lambda-\lambda p)^{n-m}\mathrm{e}^{-\lambda}}{m!(n-m)!}\cdot\frac{n!}{\lambda^n\mathrm{e}^{-\lambda}}$$

$$=\binom{n}{m}p^m(1-p)^{n-m},\quad m=0,1,2,\cdots,n$$

(4) $P\{\nu-\mu=k\}=\sum\limits_{j=0}^{\infty}P\{\mu=j,\nu=k+j\}$

$$=\sum_{j=0}^{\infty}P\{\nu=k+j\}P\{\mu=j|\nu=k+j\}$$

$$=\sum_{j=0}^{\infty}\frac{\lambda^{k+j}}{(k+j)!}\mathrm{e}^{-\lambda}\frac{(k+j)!}{j!\,k!}p^j(1-p)^k$$

$$=\frac{(1-p)^k\lambda^k}{k!}\mathrm{e}^{-\lambda+\lambda p}\sum_{j=0}^{\infty}\frac{\lambda^jp^j}{j!}\mathrm{e}^{-\lambda p}$$

$$=\frac{(\lambda-\lambda p)^k}{k!}\mathrm{e}^{-(\lambda-\lambda p)},\quad k=0,1,2,\cdots$$

【评注】很有名的一个模型, 习题二 43 题为其实例之一, 后面还要再讨论.

18. 设二维随机变量 (ξ,η) 的联合密度为

$$p(x,y)=\frac{1}{\Gamma(k_1)\Gamma(k_2)}x^{k_1-1}(y-x)^{k_2-1}\mathrm{e}^{-y}$$

$k_1>0$, $k_2>0$, $0<x\leqslant y<\infty$. 试求 ξ 与 η 的边际分布密度.

提示: 求边际密度, 注意积分限的选取.

答 ξ 的边际分布密度

$$p_\xi(x)=\frac{x^{k_1-1}}{\Gamma(k_1)}\mathrm{e}^{-x},\quad x>0$$

η 的边际分布密度

$$p_\eta(y)=\frac{1}{\Gamma(k_1+k_2)}y^{k_1+k_2-1}\mathrm{e}^{-y},\quad y>0$$

【评注】当 k_1 与 k_2 为正整数时, ξ 表示泊松过程中第 k_1 个来到时刻, η 表示第 k_1+k_2 来到时刻, 这就是本题的实际背景.

*19. 试证 $p(x,y)=K\mathrm{e}^{-(ax^2+2bxy+cy^2)}$ 为密度函数的充要条件为 $a>0$, $c>0$, $b^2-ac<0$, $K=\dfrac{1}{\pi}\sqrt{ac-b^2}$.

证 先证必要性. 在积分

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\,\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K\mathrm{e}^{-a\left(x+\frac{b}{a}y\right)^2}\mathrm{e}^{-\frac{ac-b^2}{a}y^2}\,\mathrm{d}x\mathrm{d}y$$

中, 令 $u=x+\dfrac{b}{a}y$, $v=y$, 解得 $y=v$, $x=u-\dfrac{b}{a}v$, 变换的雅可比行列式 $|J|=1$. 如果 $p(x,y)$ 是一个密度函数, 则应有

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\,\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{\infty}K\mathrm{e}^{-au^2}\,\mathrm{d}u\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{ac-b^2}{a}v^2}\,\mathrm{d}v=1$$

这意味着上述积分应收敛, 因而必然要求 $a>0$, $\dfrac{ac-b^2}{a}>0$, 由此推得 $ac-b^2>0$ 以及 $c>0$.

由等式 $\displaystyle\int_{-\infty}^{\infty}\mathrm{e}^{-at^2}\mathrm{d}t=\frac{\sqrt{\pi}}{\sqrt{a}}$ 及

$$\int_{-\infty}^{\infty}K\mathrm{e}^{-au^2}\mathrm{d}u\int_{-\infty}^{\infty}\mathrm{e}^{-\frac{ac-b^2}{a}v^2}\mathrm{d}v=K\cdot\frac{\sqrt{\pi}}{\sqrt{a}}\cdot\frac{\sqrt{a}}{\sqrt{ac-b^2}}\sqrt{\pi}=1$$

推得 $K=\dfrac{\sqrt{ac-b^2}}{\pi}$, 从而证得必要性.

为证明充分性, 必须证明在所列条件下, $p(x,y)\geqslant 0$ 及

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\,\mathrm{d}x\mathrm{d}y=1$$

前者是显然的, 后者只要把上述证明步骤逆推, 即可证得.

【评注】事实上 $ax^2+2bxy+cy^2=\begin{pmatrix}x & y\end{pmatrix}\begin{pmatrix}a & b\\ b & c\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}$, 条件 $a>0$, $c>0$, $ac-b^2>0$, 保证了矩阵 $\begin{pmatrix}a & b\\ b & c\end{pmatrix}$ 是正定矩阵.

在 $K\mathrm{e}^{-Q(x,y)}$ 中, $Q(x,y)$ 为正定二次型, 才能构成密度, 且必为正态分布.

20. (1) 若 (ξ,η) 的联合密度函数为

$$f(x,y)=\begin{cases}4xy, & 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1\\ 0, & \text{其他}\end{cases}$$

问 ξ 与 η 是否相互独立?

(2) 若 (ξ,η) 的联合密度函数为

$$g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y,\ 0\leqslant y\leqslant 1\\ 0, & \text{其他}\end{cases}$$

问 ξ 与 η 是否相互独立?

答 (1) $f_\xi(x)=2x$, $0\leqslant x\leqslant 1$, $f_\eta(y)=2y$, $0\leqslant y\leqslant 1$, 易知 ξ 与 η 相互独立.

(2) $g_\xi(x)=4x(1-x^2)$, $0\leqslant x\leqslant 1$, $g_\eta(y)=4y^3$, $0\leqslant y\leqslant 1$, $g(x,y)\neq g_\xi(x)g_\eta(y)$, 所以 ξ 与 η 不独立.

【评注】(2) 为反例之一 (关于独立性的密度分离). 表面上 $g(x,y)$ 已写成分离形式, 但问题出在取值范围上, 请注意 $g(x,y)=8xy\cdot\mathbf{1}_{(0\leqslant x\leqslant y\leqslant 1)}$, 因此独立性要求的分离 (3.2.25) 无法实现.

21. 若 ξ,η 相互独立且皆以概率 $\dfrac{1}{2}$ 取值 $+1$ 及 -1, 令 $\zeta=\xi\eta$, 试证 ξ,η,ζ 两两独立但不相互独立.

证 $P\{\zeta=1\}=P\{\xi=1,\eta=1\}+P\{\xi=-1,\eta=-1\}=\dfrac{1}{2}$

$$P\{\zeta=-1\}=P\{\xi=1,\eta=-1\}+P\{\xi=-1,\eta=1\}=\frac{1}{2}$$

$$P\{\xi=1,\zeta=1\}=P\{\xi=1,\eta=1\}=\frac{1}{4}=P\{\xi=1\}P\{\zeta=1\}$$

同理可得

$$P\{\xi=1,\zeta=-1\}=P\{\xi=1\}P\{\zeta=-1\}$$

$$P\{\xi=-1,\zeta=1\}=P\{\xi=-1\}P\{\zeta=1\}$$

$$P\{\xi=-1,\zeta=-1\}=P\{\xi=-1\}P\{\zeta=-1\}$$

故 ξ 与 ζ 相互独立.

类似可证 η 与 ζ 相互独立. 即有 ξ,η,ζ 两两独立.

但

$$P\{\xi=1,\eta=1,\zeta=1\}=P\{\xi=1,\eta=1\}=\frac{1}{4}$$

$$P\{\xi=1\}P\{\eta=1\}P\{\zeta=1\}=\frac{1}{8}$$

所以 ξ,η,ζ 不相互独立.

【评注】反例之二 (关于两两独立与相互独立). 两两独立不意味着相互独立, 随机事件如此, 随机变量也如此.

两两独立而不相互独立之例, 首推伯恩斯坦反例, 已在《概率论基础》第二章 §2 介绍过.

另一个有名的例子是: 投硬币, A 表示第一次出正面, B 表示第二次出正面, C 表示两次同为正面或同为反面, 则 A,B,C 两两独立而不相互独立.

还有一个常用的例子是: 掷两颗骰子, A 表示第一颗出奇数, B 表示第二颗出奇数, C 表示二颗之和为奇数, 则 A,B,C 两两独立而不相互独立.

事实上, 这三个例子都可以化为本习题, 即它们是等价的.

22. 设 (ξ,η) 具有联合密度函数

$$p(x,y)=\begin{cases}\dfrac{1+xy}{4}, & |x|<1,\ |y|<1\\ 0, & \text{其他}\end{cases}$$

试证 ξ 与 η 不独立, 但 ξ^2 与 η^2 是相互独立的.

证 (1) $f_\xi(x)=\dfrac{1}{2},\quad |x|<1,\quad g_\eta(y)=\dfrac{1}{2},\quad |y|<1$

因 ξ 与 η 的联合密度函数不能分离变量, 即 $p(x,y)$ 不能写成为 $f_\xi(x)g_\eta(y)$ 的形式, 所以 ξ,η 不独立.

(2) 对 $0 \leqslant u \leqslant 1$,

$$P\{\xi^2 < u\} = \int\limits_{x^2<u} f_\xi(x)\,\mathrm{d}x = \int\limits_{x^2<u} \frac{1}{2}\,\mathrm{d}x = \sqrt{u}$$

同理对 $0 \leqslant v \leqslant 1$,

$$P\{\eta^2 < v\} = \sqrt{v}$$

而对 $0 \leqslant u \leqslant 1, 0 \leqslant v \leqslant 1$,

$$P\{\xi^2 < u,\ \eta^2 < v\} = \iint\limits_{\substack{x^2<u \\ y^2<v}} p(x,y)\,\mathrm{d}x\mathrm{d}y = \sqrt{u}\sqrt{v}$$

因 ξ^2 与 η^2 的联合分布函数能写成边际分布乘积的形式, 所以 ξ^2 与 η^2 相互独立.

【评注】反例之三 (关于随机变量函数的独立性). 是否与定理 3.3.2 矛盾? ——不, 因为 ξ 与 η 不是 ξ^2 与 η^2 的单值函数.

23. 若每次试验中出现 A_1, A_2, A_3 的概率分别为 p_1, p_2 及 p_3, 而且 $p_1+p_2+p_3=1$, 共进行 n 次独立试验, 以 μ_1, μ_2, μ_3 分别记 A_1, A_2, A_3 出现的次数, 试求:

(1) (μ_1, μ_2) 的联合概率分布;

(2) μ_1 的概率分布;

(3) $P\{\mu_2 = k_2 \,|\, \mu_1 = k_1\}$.

答 (1)
$$\begin{aligned} P\{\mu_1=k_1,\mu_2=k_2\} &= P\{\mu_1=k_1,\mu_2=k_2,\mu_3=n-k_1-k_2\} \\ &= \frac{n!}{k_1!k_2!(n-k_1-k_2)!}\,p_1^{k_1}p_2^{k_2}p_3^{n-k_1-k_2} \end{aligned}$$

(2)
$$\begin{aligned} P\{\mu_1=k_1\} &= \sum_{k_2=0}^{n-k_1} P\{\mu_1=k_1,\mu_2=k_2\} \\ &= \frac{n!}{k_1!(n-k_1)!}p_1^{k_1}(1-p_1)^{n-k_1} \end{aligned}$$

(3)
$$P\{\mu_2=k_2|\mu_1=k_1\} = \frac{P\{\mu_1=k_1,\mu_2=k_2\}}{P\{\mu_1=k_1\}}$$

$$= \frac{P\{\mu_1 = k_1, \mu_2 = k_2, \mu_3 = n - k_1 - k_2\}}{P\{\mu_1 = k_1\}}$$

$$= \frac{n!\, p_1^{k_1} p_2^{k_2} p_3^{n-k_1-k_2}}{k_1! k_2! (n - k_1 - k_2)!} \cdot \frac{k_1!(n-k_1)!}{n!\, p_1^{k_1} (1-p_1)^{n-k_1}}$$

$$= \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \left(\frac{p_2}{p_2+p_3}\right)^{k_2} \left(\frac{p_3}{p_2+p_3}\right)^{n-k_1-k_2}$$

【评注】以 $r = 3$ 为例讨论多项分布. 模型: 推广的伯努利试验 (多色放回摸球模型); $r = 2$ 化为二项分布; 边际分布仍为多项分布; 条件分布仍为多项分布.

24. 袋中装 i 号球 N_i 只, $i = 1, 2, 3$, $N_1 + N_2 + N_3 = N$, 从中随机摸出 n 只, 若以 μ_1, μ_2, μ_3 分别记 1, 2, 3 号球出现的次数, 试求:

(1) (μ_1, μ_2) 的联合概率分布;

(2) μ_1 的概率分布;

(3) $P\{\mu_2 = n_2 \,|\, \mu_1 = n_1\}$.

答

(1) $$P\{\mu_1 = n_1, \mu_2 = n_2\} = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\binom{N - N_1 - N_2}{n - n_1 - n_2}}{\binom{N}{n}}$$

(2) $$P\{\mu_1 = n_1\} = \sum_{n_2=0}^{n-n_1} P\{\mu_1 = n_1, \mu_2 = n_2\} = \frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}}$$

(3) $$P\{\mu_2 = n_2 | \mu_1 = n_1\} = \frac{\binom{N_2}{n_2}\binom{N_3}{n_3}}{\binom{N - N_1}{n - n_1}}$$

其中 $n_3 = n - n_1 - n_2$.

【评注】与上题对照讨论 $r = 3$ 的多元超几何分布. 模型: 多色不放回摸球模型; $r = 2$ 化为超几何分布; 边际分布仍为多元超

几何分布; 条件分布仍为多元超几何分布.

以上两题虽对 $r=3$ 命题, 意在对任意 $r\,(r\geqslant 3)$ 建立普遍结论. 这样做的理由是避免繁复计算, 节省做题时间.

25. 若 ξ_1 与 ξ_2 是独立随机变量, 且 $\xi_1\sim B(n_1,p)$, $\xi_2\sim B(n_2,p)$, 试直接证明

(1) $\xi_1+\xi_2\sim B(n_1+n_2,p)$;

(2) $P\{\xi_1=k\,|\,\xi_1+\xi_2=n\}=\dfrac{\dbinom{n_1}{k}\dbinom{n_2}{n-k}}{\dbinom{n_1+n_2}{n}}$.

证 (1)

$$
\begin{aligned}
P\{\xi_1+\xi_2=k\}&=\sum_{j=0}^{k}P\{\xi_1=j\}P\{\xi_2=k-j\}\\
&=\sum_{j=0}^{k}\binom{n_1}{j}p^j(1-p)^{n_1-j}\binom{n_2}{k-j}p^{k-j}(1-p)^{n_2-k+j}\\
&=\binom{n_1+n_2}{k}p^k(1-p)^{n_1+n_2-k}
\end{aligned}
$$

所以得证 $\xi_1+\xi_2\sim B(n_1+n_2,p)$;

(2)
$$
\begin{aligned}
P\{\xi_1=k\,|\,\xi_1+\xi_2=n\}&=\frac{P\{\xi_1=k,\xi_1+\xi_2=n\}}{P\{\xi_1+\xi_2=n\}}\\
&=\frac{P\{\xi_1=k\}P\{\xi_2=n-k\}}{P\{\xi_1+\xi_2=n\}}\\
&=\frac{\dbinom{n_1}{k}p^k(1-p)^{n_1-k}\dbinom{n_2}{n-k}p^{n-k}(1-p)^{n_2-n+k}}{\dbinom{n_1+n_2}{n}p^n(1-p)^{n_1+n_2-n}}\\
&=\frac{\dbinom{n_1}{k}\dbinom{n_2}{n-k}}{\dbinom{n_1+n_2}{n}}
\end{aligned}
$$

【评注】(1) 讨论二项分布的一个重要性质——再生性. 把一枚硬币投了 n_1 次之后再投 n_2 次, 关心出现几次正面, 答案很明显.

(2) 条件分布的答案很有趣, 居然是超几何分布.

26. 若 ξ_1 与 ξ_2 是独立随机变量, 均服从泊松分布, 参数分别为 λ_1 及 λ_2, 试直接证明:

(1) $\xi_1+\xi_2$ 具有泊松分布, 参数为 $\lambda_1+\lambda_2$;

(2) $P\{\xi_1=k\,|\,\xi_1+\xi_2=n\}=\binom{n}{k}\left(\dfrac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\dfrac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$.

证 (证明留给读者)

【评注】与上题对照, 讨论泊松分布.

(1) 再生性也成立, 但没有上题直观: 客车流加货车流还是泊松流?!

(2) 条件分布对应地变成二项分布.

27. 设 ξ 的密度函数为 $p(x)$, 求下列随机变量的分布密度函数: (1) $\eta=\xi^{-1}$; (2) $\eta=\tan\xi$; (3) $\eta=|\xi|$.

提示: 用直接法或变换法均可.

答 (1) $p(y^{-1})y^{-2}$, $y\neq 0$;

(2) $\dfrac{1}{1+y^2}\displaystyle\sum_{n=-\infty}^{\infty}p(n\pi+\arctan y)$;

(3) $p(-y)+p(y)$, $y\geqslant 0$.

【评注】求连续型随机变量函数的分布.

中等难度, 不止一种做法.

28. 设 ξ 与 η 相互独立且服从同一几何分布, 令 $\zeta=\max(\xi,\eta)$, 试求 (1) (ζ,ξ) 的联合概率分布; (2) ζ 的概率分布; (3) ξ 关于 ζ 的条件概率分布.

解 设 $P\{\xi=i\}=q^{i-1}p$, $i=1,2,\cdots$,

(1) 当 $i<k$, $P\{\zeta=k,\xi=i\}=P\{\eta=k,\xi=i\}=p^2q^{i+k-2}$;

当 $i=k$, $P\{\zeta=k,\xi=i\}=\sum\limits_{j=1}^{k}P\{\eta=j,\xi=k\}=pq^{k-1}(1-q^k)$;

当 $i>k$, $P\{\zeta=k,\xi=i\}=0$;

(2)
$$P\{\zeta=k\}=\sum_{i=1}^{k-1}P\{\xi=i,\eta=k\}+\sum_{j=1}^{k}P\{\xi=k,\eta=j\}$$
$$=\sum_{i=1}^{k-1}p^2q^{i+k-2}+pq^{k-1}(1-q^k)$$
$$=pq^{k-1}(2-q^{k-1}-q^k)$$

(3)
$$P\{\xi=i|\zeta=k\}=\frac{P\{\xi=i,\zeta=k\}}{P\{\zeta=k\}}$$
$$=\begin{cases}\dfrac{pq^{i-1}}{2-q^{k-1}-q^k}, & i<k\\ \dfrac{1-q^k}{2-q^{k-1}-q^k}, & i=k\\ 0, & i>k\end{cases}$$

【评注】求离散型随机变量函数的分布.

29. 若 ξ,η 为相互独立的分别服从 $[0,1]$ 均匀分布的随机变量, 试求 $\zeta=\xi+\eta$ 的分布密度函数.

解法一 (卷积公式) ξ 与 η 的密度函数为

$$p_\xi(x)=p_\eta(x)=\begin{cases}1, & 0\leqslant x\leqslant 1\\ 0, & 其他\end{cases}$$

由卷积公式 (3.3.20) 及独立性得 $\zeta=\xi+\eta$ 的分布密度函数为

$$p_\zeta(x)=\int_{-\infty}^{\infty}p_\xi(t)p_\eta(x-t)\,\mathrm{d}t$$
$$=\int\limits_{\substack{0\leqslant t\leqslant 1\\ 0\leqslant x-t\leqslant 1}}1\times 1\,\mathrm{d}t=\int\limits_{\substack{0\leqslant t\leqslant x,0\leqslant x\leqslant 1\\ x-1\leqslant t\leqslant 1,1\leqslant x\leqslant 2}}\mathrm{d}t$$

$$
=\begin{cases}\int_0^x \mathrm{d}t, & 0 \leqslant x \leqslant 1 \\ \int_{x-1}^1 \mathrm{d}t, & 1 \leqslant x \leqslant 2 \\ 0, & \text{其他}\end{cases}
$$

$$
=\begin{cases}x, & 0 \leqslant x \leqslant 1 \\ 2-x, & 1 \leqslant x \leqslant 2 \\ 0, & \text{其他}\end{cases}
$$

解法二 (几何方法) 由图 3-2 可知, 若记

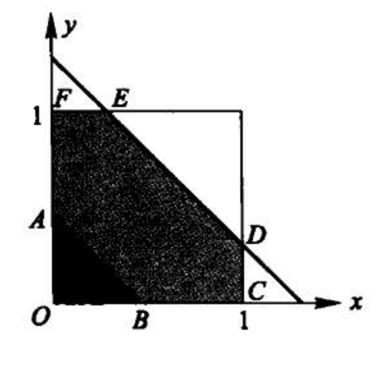


图 3-2

$$
F_\zeta(z)=P\{\xi+\eta<z\}=\iint\limits_{x+y<z} p(x,y)\mathrm{d}x\mathrm{d}y
$$

则当 $z \leqslant 0$ 时, $F_\zeta(z)=0$;

当 $0<z\leqslant 1$ 时, $F_\zeta(z)=\iint\limits_{0\leqslant x+y<z} \mathrm{d}x\mathrm{d}y=\dfrac{z^2}{2}$

(即 $\triangle AOB$ 之面积);

当 $1<z\leqslant 2$ 时,

$$
F_\zeta(z)=\iint\limits_{\substack{0\leqslant x+y<z\\0\leqslant x\leqslant 1\\0\leqslant y\leqslant 1}} \mathrm{d}x\mathrm{d}y=1-\frac{(2-z)^2}{2}=-\frac{z^2}{2}+2z-1
$$

(即多边形 $OCDEF$ 之面积);

当 $z>2$ 时, $F_\zeta(z)=\iint\limits_{\substack{0\leqslant x\leqslant 1\\0\leqslant y\leqslant 1}}=1$

于是, $\zeta=\xi+\eta$ 的密度函数为

$$
p_\zeta(x)=\begin{cases}x, & 0 \leqslant x \leqslant 1 \\ 2-x, & 1 \leqslant x \leqslant 2 \\ 0, & \text{其他}\end{cases}
$$

【评注】解法一是卷积公式 (3.3.20) 的用法举例，该公式使用中的主要困难之一是确定积分上下限. 几何方法则更直观，更简便，当然实质是一致的.

这个分布称为三角形分布，亦称辛普森分布.

辛普森 (1710—1761) 在数学中因数值计算而著名，对概率论主要贡献在误差理论，是概率论中最早研究连续分布的学者.

30. 在 $(0, a)$ 线段上随机投掷两点，试求两点间距离的分布函数.

解 设在 $[0,a]$ 内任意投掷两点，它们的坐标记为 (ξ_1,ξ_2)，则 (ξ_1,ξ_2) 的联合密度函数为

$$p(t_1,t_2)=\begin{cases}\dfrac{1}{a^2}, & 0\leqslant t_1\leqslant a,\ 0\leqslant t_2\leqslant a\\ 0, & \text{其他}\end{cases}$$

设两点距离为 $\eta=|\xi_1-\xi_2|$，其分布函数记为 $F_\eta(x)$，从图 3-3 可看出满足要求的为图中的阴影部分，故

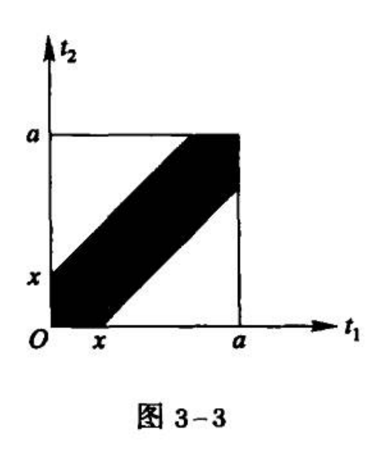


图 3-3

$$\begin{aligned}F_\eta(x)&=P\{|\xi_1-\xi_2|<x\}\\&=\iint\limits_{\substack{-x<t_1-t_2<x\\0<t_1<a\\0<t_2<a}}p(t_1,t_2)\,\mathrm{d}t_1\mathrm{d}t_2\\&=\begin{cases}0, & x\leqslant 0\\ \dfrac{a^2-(a-x)^2}{a^2}, & 0<x\leqslant a\\ 1, & x>a\end{cases}\end{aligned}$$

【评注】从解题的几何图示可明显看出本题与会面问题的联系.

31. 若 ξ 与 η 相互独立，均服从参数为 $\dfrac{1}{\alpha}$ 的指数分布 $\mathrm{Exp}\left(\dfrac{1}{\alpha}\right)$,

试证 $\zeta=\xi-\eta$ 服从拉普拉斯分布:

$$p(x)=\frac{1}{2\alpha}\mathrm{e}^{-\frac{|x|}{\alpha}},\quad -\infty<x<\infty$$

证 ξ 与 η 的密度函数为:

$$q(t)=\frac{1}{\alpha}\mathrm{e}^{-\frac{t}{\alpha}},\quad t>0$$

记 ζ 的分布函数为 $F(x)=P\{\xi-\eta<x\}$, 则
当 $x>0$ 时, 参看图 3-4

$$\begin{aligned}
F(x)&=\int_0^{+\infty}\mathrm{d}t_2\int_0^{t_2+x}\frac{1}{\alpha^2}\mathrm{e}^{-\frac{t_1+t_2}{\alpha}}\,\mathrm{d}t_1\\
&=\int_0^{+\infty}\frac{1}{\alpha}\mathrm{e}^{-\frac{t_2}{\alpha}}\left[-\mathrm{e}^{-\frac{t_1}{\alpha}}\right]_0^{t_2+x}\mathrm{d}t_2\\
&=\int_0^{+\infty}\frac{1}{\alpha}\mathrm{e}^{-\frac{t_2}{\alpha}}\left[1-\mathrm{e}^{-\frac{t_2+x}{\alpha}}\right]\mathrm{d}t_2\\
&=\int_0^{+\infty}\left[\frac{1}{\alpha}\mathrm{e}^{-\frac{t_2}{\alpha}}-\frac{1}{\alpha}\mathrm{e}^{-\frac{2t_2+x}{\alpha}}\right]\mathrm{d}t_2\\
&=-\mathrm{e}^{-\frac{t_2}{\alpha}}\Big|_0^{+\infty}+\frac{1}{2}\mathrm{e}^{-\frac{2t_2+x}{\alpha}}\Big|_0^{+\infty}\\
&=1-\frac{1}{2}\mathrm{e}^{-\frac{x}{\alpha}}
\end{aligned}$$

由此得, 当 $x>0$ 时的 ζ 的密度函数为 $p(x)=\frac{1}{2\alpha}\mathrm{e}^{-\frac{x}{\alpha}}$, $x>0$,
当 $x\leqslant 0$ 时, 参看图 3-5

$$\begin{aligned}
F(x)&=\int_{-x}^{+\infty}\mathrm{d}t_2\int_0^{t_2+x}\frac{1}{\alpha^2}\mathrm{e}^{-\frac{t_1+t_2}{\alpha}}\,\mathrm{d}t_1\\
&=\int_{-x}^{+\infty}\left[\frac{1}{\alpha}\mathrm{e}^{-\frac{t_2}{\alpha}}-\frac{1}{\alpha}\mathrm{e}^{-\frac{2t_2+x}{\alpha}}\right]\mathrm{d}t_2\\
&=-\mathrm{e}^{-\frac{t_2}{\alpha}}\Big|_{-x}^{+\infty}+\frac{1}{2}\mathrm{e}^{-\frac{2t_2+x}{\alpha}}\Big|_{-x}^{+\infty}
\end{aligned}$$

$$=\mathrm{e}^{\frac{x}{\alpha}}-\frac{1}{2}\mathrm{e}^{\frac{x}{\alpha}}$$

$$=\frac{1}{2}\mathrm{e}^{\frac{x}{\alpha}}$$

于是 ζ 的密度函数在 $x\leqslant 0$ 时为 $p(x)=\dfrac{1}{2\alpha}\mathrm{e}^{\frac{x}{\alpha}}$, $x\leqslant 0$.

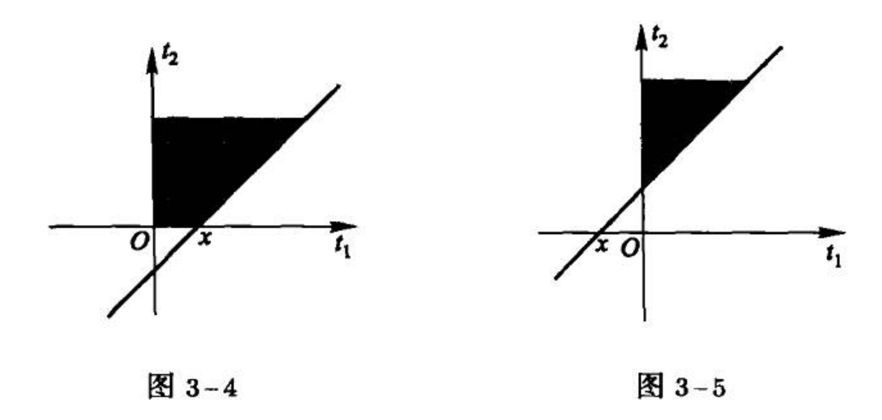


图 3-4　　　　图 3-5

所以得 ζ 的密度函数为

$$p(x)=\frac{1}{2\alpha}\mathrm{e}^{-\frac{|x|}{\alpha}}$$

【评注】拉普拉斯分布的出处之一. 也可用差的公式 (其推导类似于和的公式) 计算. 困难之一仍在积分限的确定.

经验表明, 使用和或差的公式求密度函数, 最容易出错的是积分上下限的选定.

32. 若 ξ 与 η 相互独立, 分别服从 $N(0,1)$, 试证 $\psi=\dfrac{\xi}{\eta}$ 服从柯西分布.

证　ψ 的密度函数为

$$p(x)=\int_{-\infty}^{+\infty}|z|\varphi(zx)\varphi(z)\,\mathrm{d}z$$

$$=\int_{-\infty}^{+\infty}|z|\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2x^2}{2}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\,\mathrm{d}z$$

$$= 2\,\frac{1}{2\pi}\int_0^{+\infty} z\,\mathrm{e}^{-\frac{(1+x^2)z^2}{2}}\,\mathrm{d}z$$

$$= -\frac{1}{\pi}\,\frac{1}{1+x^2}\,\mathrm{e}^{-\frac{(1+x^2)z^2}{2}}\Big|_0^{+\infty}$$

$$= \frac{1}{\pi}\cdot\frac{1}{1+x^2}$$

【评注】柯西分布的出处之一. 商的公式 (3.3.23) 的应用例子.

33. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立, 且皆服从指数分布, 参数分别为 $\lambda_1,\lambda_2,\cdots,\lambda_n$, 试求 $\eta=\min(\xi_1,\xi_2,\cdots,\xi_n)$ 的分布.

解 指数分布的密度函数 $p(x)=\lambda\,\mathrm{e}^{-\lambda x}$, $x>0$, 其分布函数为 $F(x)=1-\mathrm{e}^{-\lambda x}$, $x>0$.

η 的分布函数为

$$F_\eta(x)=1-\prod_{i=1}^{n}\big(1-F_i(x)\big)=1-\mathrm{e}^{-(\lambda_1+\cdots+\lambda_n)x},\ x>0$$

η 的密度函数:

$$p_\eta(x)=(\lambda_1+\cdots+\lambda_n)\,\mathrm{e}^{-(\lambda_1+\cdots+\lambda_n)x},\ x>0$$

即 η 服从参数为 $\lambda_1+\cdots+\lambda_n$ 的指数分布.

【评注】指数分布的重要性质之一. 事实上与习题三 26 题关系密切, 也见下题.

34. 通称下列分布函数为**韦布尔分布**:

$$F(x)=\begin{cases}1-\mathrm{e}^{-\lambda x^\alpha}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

这是韦布尔 (Weibull) 在研究金属材料的疲劳寿命中导出的, 在可靠性研究中有广泛应用.

若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立, 相同分布, 并以 ξ_1^* 记它们的最小值. (1) 当 $\xi_i\sim F(x)$ 时, 试求 ξ_1^* 的分布函数; (2) 若 $\xi_1^*\sim F(x)$, 试导出 ξ_i 的分布函数.

链条的寿命取决于最弱环节, 试说明上述概率结论的实际含意.

提示: 回忆《概率论基础》中最小值公式 (3.3.26).

答 (1) $1-\mathrm{e}^{-n\lambda x^{\alpha}}$, $x>0$;

(2) $1-\mathrm{e}^{-\frac{\lambda}{n}x^{\alpha}}$, $x>0$.

【评注】韦布尔分布亦是概率论中有名的分布, 当形状参数 $\alpha=1$ 时化为指数分布. 它在可靠性研究中常用作寿命分布, 理由之一可从如下实例中悟出: 链条的寿命与构成它的单环的疲劳寿命相关, 若其一服从韦布尔分布, 则另一个也是.

35. 若 (ξ,η) 服从二元正态分布, 参数 μ_1, μ_2, σ_1^2, σ_2^2, ρ, 以 $D(\lambda)$ 记下面椭圆的内部:

$$\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}=\lambda^2$$

试求 $P\{(\xi,\eta)\in D(\lambda)\}$.

解 所求概率为

$$\begin{aligned}P\{(\xi,\eta)\in D(\lambda)\}&=\iint\limits_{(x,y)\in D(\lambda)}p(x,y)\,\mathrm{d}x\mathrm{d}y\\&=\iint\limits_{\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\leqslant\lambda^2}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\\&\quad\cdot\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right.\right.\\&\quad\left.\left.+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right\}\mathrm{d}x\mathrm{d}y\end{aligned}$$

令 $x=\mu_1+\sigma_1 r\cos\theta$, $y=\mu_2+\sigma_2 r\sin\theta$, $|J|=\sigma_1\sigma_2 r$,

$$\frac{(x-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}$$

$$= r^2\cos^2\theta - 2\rho r^2\sin\theta\cos\theta + r^2\sin^2\theta$$

$$\triangleq r^2 h(\theta)$$

则

$$P\{(\xi,\eta)\in D(\lambda)\}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_0^{2\pi}\mathrm{d}\theta\int_0^{\frac{\lambda}{\sqrt{h(\theta)}}}\mathrm{e}^{-\frac{r^2h(\theta)}{2(1-\rho^2)}}\sigma_1\sigma_2 r\,\mathrm{d}r$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_0^{2\pi}\mathrm{d}\theta\int_0^{\frac{\lambda}{\sqrt{h(\theta)}}}\left(-\frac{1-\rho^2}{h(\theta)}\right)\mathrm{d}\,\mathrm{e}^{-\frac{r^2h(\theta)}{2(1-\rho^2)}}$$

$$= \frac{\sqrt{1-\rho^2}}{2\pi}\int_0^{2\pi}\left(1-\mathrm{e}^{-\frac{\lambda^2}{2(1-\rho^2)}}\right)\frac{\mathrm{d}\theta}{h(\theta)}$$

$$= \left(1-\mathrm{e}^{-\frac{\lambda^2}{2(1-\rho^2)}}\right)\frac{\sqrt{1-\rho^2}}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\theta}{h(\theta)}$$

令 $\lambda\to\infty$, 得 $\dfrac{\sqrt{1-\rho^2}}{2\pi}\displaystyle\int_0^{2\pi}\dfrac{\mathrm{d}\theta}{h(\theta)}=1$, 故

$$P\{(\xi,\eta)\in D(\lambda)\} = 1-\mathrm{e}^{-\frac{\lambda^2}{2(1-\rho^2)}}$$

【评注】计算了二元正态分布在等概率椭圆内的分布值, 也可以看作瑞利分布 (3.3.41) 的推广, 或至少是另一种推导.

36. 若气体分子的速度是随机向量 $V=(X,Y,Z)$, 各分量相互独立, 且均服从 $N(0,\sigma^2)$, 试证 $S=\sqrt{X^2+Y^2+Z^2}$ 服从**麦克斯韦分布律**:

$$p(s)=\sqrt{\frac{2}{\pi}}\,\frac{s^2}{\sigma^3}\,\exp\left(-\frac{s^2}{2\sigma^2}\right),\quad s>0$$

证 $V(X,Y,Z)$ 的概率密度为

$$p(x,y,z)=\frac{1}{\left(\sqrt{2\pi}\right)^3\sigma^3}\,\mathrm{e}^{-\frac{1}{2\sigma^2}(x^2+y^2+z^2)}$$

S 的分布函数为: $s > 0$ 时,

$$F(s) = P\{\sqrt{X^2+Y^2+Z^2} < s\}$$

$$= \iiint\limits_{\sqrt{x^2+y^2+z^2}<s} \frac{1}{(\sqrt{2\pi})^3\sigma^3} \mathrm{e}^{-\frac{1}{2\sigma^2}(x^2+y^2+z^2)} \mathrm{d}x\mathrm{d}y\mathrm{d}z$$

令 $x = \rho\cos\theta\sin\varphi$, $y = \rho\sin\theta\sin\varphi$, $z = \rho\cos\varphi$, 则

$$= \int_0^{2\pi} \mathrm{d}\theta \int_0^{\pi} \sin\varphi\,\mathrm{d}\varphi \int_0^s \frac{1}{(\sqrt{2\pi})^3\sigma^3} \mathrm{e}^{-\frac{1}{2\sigma^2}\rho^2} \rho^2\,\mathrm{d}\rho$$

$$= 2\pi\cdot 2\cdot \int_0^s \frac{1}{(\sqrt{2\pi})^3\sigma^3} \mathrm{e}^{-\frac{\rho^2}{2\sigma^2}} \rho^2\mathrm{d}\rho$$

对 s 求导, 则得 S 的密度函数为

$$p(s) = \sqrt{\frac{2}{\pi}}\,\frac{s^2}{\sigma^3}\,\exp\left(-\frac{s^2}{2\sigma^2}\right), \quad s > 0$$

【评注】19 世纪后半叶, 麦克斯韦开创的理想气体统计物理学从概率论借助工具极大地发展了物理学, 反过来也大大地促进了概率论的发展.

以上几题, 我们趁机引进不少重要分布.

*37. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立, 均服从 $N(0,1)$, 试用 (3.3.18) 式, 化为 n 维极坐标, 证明 $\eta = \xi_1^2+\xi_2^2+\cdots+\xi_n^2$ 服从 χ^2 分布.

证 对 $x > 0$,

$$F(x) = P\{\eta < x\}$$

$$= P\{\xi_1^2+\xi_2^2+\cdots+\xi_n^2 < x\}$$

$$= \int\cdots\int\limits_{x_1^2+\cdots+x_n^2<x} \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{x_1^2+\cdots+x_n^2}{2}\right\} \mathrm{d}x_1\cdots\mathrm{d}x_n$$

作变换

$$\begin{cases} x_1 &= \rho\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-1} \\ x_2 &= \rho\cos\theta_1\sin\theta_2\cdots\sin\theta_{n-1} \\ & \cdots\cdots\cdots\cdots \\ x_{n-1} &= \rho\cos\theta_{n-2}\sin\theta_{n-1} \\ x_n &= \rho\cos\theta_{n-1} \end{cases}$$

此变换的雅可比行列式为

$$J = \frac{\partial(x_1,\cdots,x_n)}{\partial(\rho,\theta_1,\cdots,\theta_{n-1})} = \rho^{n-1}D(\theta_1,\cdots,\theta_{n-1})$$

其中 $D(\theta_1,\cdots,\theta_{n-1})$ 只是 $\theta_1,\cdots,\theta_{n-1}$ 的函数, 与 ρ 无关, 于是

$$\begin{aligned} F(x) &= \frac{1}{(2\pi)^{n/2}}\int_0^{\sqrt{x}} \mathrm{e}^{-\frac{\rho^2}{2}}\rho^{n-1}\mathrm{d}\rho\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi}D(\theta_1,\cdots,\theta_{n-1})\mathrm{d}\theta_1\cdots\mathrm{d}\theta_{n-1} \\ &= C_n\int_0^{\sqrt{x}}\mathrm{e}^{-\frac{\rho^2}{2}}\rho^{n-1}\mathrm{d}\rho \end{aligned}$$

其中 $C_n = \frac{1}{(2\pi)^{n/2}}\int_0^{\pi}\cdots\int_0^{\pi}\int_0^{2\pi}D(\theta_1,\cdots,\theta_{n-1})\mathrm{d}\theta_1\cdots\mathrm{d}\theta_{n-1}$. 再令 $\rho=\sqrt{t}$,

$$F(x) = C_n\cdot\frac{1}{2}\int_0^{x}\mathrm{e}^{-\frac{t}{2}}t^{\frac{n}{2}-1}\mathrm{d}t$$

令 $x\to+\infty$,

$$\begin{aligned} 1 &= C_n\cdot\frac{1}{2}\int_0^{+\infty}\mathrm{e}^{-\frac{t}{2}}t^{\frac{n}{2}-1}\mathrm{d}t \\ &= C_n\cdot\frac{1}{2}\int_0^{+\infty}\mathrm{e}^{-u}u^{\frac{n}{2}-1}2^{\frac{n}{2}}\mathrm{d}u \\ &= C_n\cdot 2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right) \end{aligned}$$

故 $C_n = \dfrac{1}{2^{\frac{n}{2}-1}\Gamma\left(\dfrac{n}{2}\right)}$, 于是得 η 的分布函数

$$F(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}\int_0^x \mathrm{e}^{-\frac{t}{2}} t^{\frac{n}{2}-1}\mathrm{d}t, \quad x > 0$$

两边关于 x 求导, 就得 η 的概率密度函数

$$p(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}\mathrm{e}^{-\frac{x}{2}} x^{\frac{n}{2}-1}, \quad x > 0$$

即 χ^2 分布的密度函数.

【评注】χ^2 分布是数理统计中最重要的分布, 记住它是 n 个独立标准正态变量平方和的分布.

35 题算 2 维积分, 36 题算 3 维积分, 37 题算 n 维积分, 概率论学科 (特别是数理统计) 是使用高维积分最多的学科之一.

38. 若 ξ 与 η 相互独立, 且分别服从 $\Gamma(r_1,\lambda)$ 及 $\Gamma(r_2,\lambda)$, 试求 $\alpha=\xi+\eta$ 与 $\beta=\dfrac{\xi}{\xi+\eta}$ 的联合密度函数 $q(u,v)$, 并证明:

(1) 随机变量 β 服从 $\boldsymbol{\beta}$ **分布**:

$$p(v) = \frac{\Gamma(r_1+r_2)}{\Gamma(r_1)\Gamma(r_2)} v^{r_1-1}(1-v)^{r_2-1}, \quad 0 < v < 1$$

(2) 随机变量 α 与 β 独立.

证 设 $u = x+y$, $v = \dfrac{x}{x+y}$, 则可推得 $x = uv$, $y = u(1-v)$. 由此得

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u, \ |J| = u$$

于是得

$$q(u,v) = \frac{\lambda^{r_1}}{\Gamma(r_1)} x^{r_1-1}\mathrm{e}^{-\lambda x}\frac{\lambda^{r_2}}{\Gamma(r_2)} y^{r_2-1}\mathrm{e}^{-\lambda y}|J|$$

$$= \frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)}\, u^{r_1-1}v^{r_1-1}\, u^{r_2-1}(1-v)^{r_2-1}\mathrm{e}^{-\lambda u}\cdot u$$

$$= \frac{\lambda^{r_1+r_2}}{\Gamma(r_1+r_2)}\, u^{r_1+r_2-1}\mathrm{e}^{-\lambda u}\cdot \frac{\Gamma(r_1+r_2)}{\Gamma(r_1)\Gamma(r_2)}\, v^{r_1-1}(1-v)^{r_2-1}$$

$$= p_\alpha(u)p_\beta(v)$$

可见, β 服从 $\boldsymbol{\beta}$ **分布**, 且 α 与 β 独立.

【评注】变换法应用的例子. 1. 导出统计中重要的 β 分布, 当 $r_1=r_2=1$ 时化为均匀分布; 2. 揭露了 Γ 分布的可加性; 3. 证明 α 与 β 独立. 此外应注意本题与《概率论基础》第三章 § 3 例 7 有明显关系.

39. 若 ξ,η 独立, 且均服从 $N(0,1)$, 试求 $U=\xi^2+\eta^2$ 与 $V=\dfrac{\xi}{\eta}$ 的密度函数, 并证明它们是独立的.

解 设 $u=x^2+y^2$, $v=\dfrac{x}{y}$, 则

$$J^{-1}=\frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix} 2x & 2y \\ \dfrac{1}{y} & -\dfrac{x}{y^2}\end{vmatrix}=-2(v^2+1)$$

$$|J|=\frac{1}{2(1+v^2)}$$

(U,V) 的联合密度函数为

$$q(u,v)=p(x,y)|J|\left(\begin{smallmatrix}-\infty<x<+\infty\\ y\geqslant 0\end{smallmatrix}\right)+p(x,y)|J|\left(\begin{smallmatrix}-\infty<x<+\infty\\ y<0\end{smallmatrix}\right)$$

$$=2\cdot\frac{1}{2\pi}\mathrm{e}^{-\frac{x^2+y^2}{2}}\frac{1}{2(1+v^2)}$$

$$=\frac{1}{2}\mathrm{e}^{-\frac{u}{2}}\cdot\frac{1}{\pi(1+v^2)}$$

$$=p_U(u)\cdot p_V(v)$$

于是可见, U, V 是相互独立的, 它们的密度函数分别为

$$p_U(u) = \frac{1}{2}\mathrm{e}^{-\frac{u}{2}}, \ u > 0$$

$$p_V(v) = \frac{1}{\pi(1+v^2)}, \ -\infty < v < +\infty$$

【评注】1. 变换法应用的例子. 这时计算 J^{-1} 较方便. 另一注意要点是这里的变换是二对一, 即 (x, y) 与 $(-x, -y)$ 对应同一 (u, v);

2. 与《概率论基础》第三章 § 3 例 9 关系密切, U, V 的独立性可预期;

3. $U = \xi^2 + \eta^2 \sim \chi_2^2$, 即 $\Gamma\left(1, \frac{1}{2}\right)$, 也即 $\mathrm{Exp}\left(\frac{1}{2}\right)$;

4. $V = \frac{\xi}{\eta}$ 服从柯西分布, 其推导又见《概率论基础》习题三 32 题, 且可通过上述例 9 与例 5 的定义式 (3.3.13) 导出.

40. 若 (ξ, η) 服从二元正态分布 (3.2.22), 试找出 $\xi+\eta$ 与 $\xi-\eta$ 相互独立的充要条件.

解 首先易得 $\xi+\eta$ 与 $\xi-\eta$ 相互独立同 $U = \xi+\eta-\mu_1-\mu_2$ 与 $V = \xi-\eta-\mu_1+\mu_2$ 相互独立是等价的. 再则令

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix}$$

则

$$\begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

且

$$|J| = \frac{1}{2}$$

$$q(u, v) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot |J| \cdot$$

$$\exp\left\{-\frac{1}{2}(u,v)\left[\begin{pmatrix}1&1\\1&-1\end{pmatrix}\begin{pmatrix}\sigma_1^2&\rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2&\sigma_2^2\end{pmatrix}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\right]^{-1}\begin{pmatrix}u\\v\end{pmatrix}\right\}$$

$$=\frac{1}{2}\cdot\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}(u,v)\boldsymbol{A}^{-1}\begin{pmatrix}u\\v\end{pmatrix}\right\}$$

其中

$$\boldsymbol{A}^{-1}=\left[\begin{pmatrix}1&1\\1&-1\end{pmatrix}\begin{pmatrix}\sigma_1^2&\rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2&\sigma_2^2\end{pmatrix}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\right]^{-1}$$

$$=\begin{pmatrix}\sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2&\sigma_1^2-\sigma_2^2\\\sigma_1^2-\sigma_2^2&\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2\end{pmatrix}^{-1}$$

由此可得 $\xi+\eta$ 与 $\xi-\eta$ 相互独立的充要条件是 $\sigma_1=\sigma_2$.

【评注】若用下章将要讲的多元正态分布性质, 本题有简明的解法.

41. 对二元正态密度函数

$$p(x,y)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(2x^2+y^2+2xy-22x-14y+65)\right\}$$

(1) 把它化为标准形式 (3.2.22); (2) 指出 μ_1, μ_2, σ_1, σ_2, ρ; (3) 求 $p_1(x)$; (4) 求 $p(x\,|\,y)$.

提示: 记住二元正态密度函数的性质, 做到不经太多计算就能写出 (2)(3)(4) 的答案.

答 (1)

$$p(x,y)=\frac{1}{2\pi\cdot1\cdot\sqrt{2}\cdot\sqrt{1-\frac{1}{2}}}$$

$$\cdot\exp\left\{\frac{1}{2\left(1-\frac{1}{2}\right)}\left[(x-4)^2+2\frac{1}{\sqrt{2}}\frac{(x-4)(y-3)}{1\cdot\sqrt{2}}+\frac{(y-3)^2}{2}\right]\right\}$$

(2) $\mu_1=4$, $\mu_2=3$, $\sigma_1=1$, $\sigma_2=\sqrt{2}$, $\rho=-\frac{1}{\sqrt{2}}$

(3) $p_1(x)=\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-4)^2}{2}}$

(4) $p(x|y)=\dfrac{1}{\sqrt{\pi}}\exp\left\{-\left[x-\left(\dfrac{11}{2}-\dfrac{y}{2}\right)\right]^2\right\}$

【评注】通过本题让学生熟悉二元正态分布.

42. 设 $\boldsymbol{\mu}=\mathbf{0}$, $\boldsymbol{\Sigma}^{-1}=\begin{pmatrix}7&3&2\\3&4&1\\2&1&2\end{pmatrix}$, 试写出分布密度 (3.2.12) 并求出 (ξ_1,ξ_2) 的边际密度函数.

答 (ξ_1,ξ_2,ξ_3) 的分布密度

$$p(x,y,z)=\frac{\sqrt{27}}{2\pi\sqrt{2\pi}}\cdot\exp\left\{-\frac{1}{2}\left(7x^2+4y^2+2z^2+6xy+4xz+2yz\right)\right\}$$

(ξ_1,ξ_2) 的边际密度函数

$$p(x,y)=\frac{3\sqrt{6}}{4\pi}\exp\left\{-\frac{1}{2}\left(5x^2+4xy+\frac{7}{2}y^2\right)\right\}$$

【评注】浅涉三元正态.

**43. 设 ξ 与 η 是相互独立相同分布的随机变量, 其密度函数不等于 0 且有二阶导数, 试证若 $\xi+\eta$ 与 $\xi-\eta$ 相互独立, 则随机变量 $\xi,\eta,\xi+\eta,\xi-\eta$ 均服从正态分布.

证 设 ξ 的密度函数为 $p(x)$, $U=\xi+\eta$, $V=\xi-\eta$ 的密度函数分别为 $f(u)$, $g(v)$, 则

$$p\left(\frac{u+v}{2}\right)p\left(\frac{u-v}{2}\right)\frac{1}{2}=f(u)g(v)$$

$$\ln p\left(\frac{u+v}{2}\right)+\ln p\left(\frac{u-v}{2}\right)-\ln 2=\ln f(u)+\ln g(v)$$

记 $h(z)=\ln p(z)$, 则

$$h\left(\frac{u+v}{2}\right)+h\left(\frac{u-v}{2}\right)-\ln 2=\ln f(u)+\ln g(v)$$

等式两边关于 u 求偏导得

$$\frac{1}{2}h'\left(\frac{u+v}{2}\right)+\frac{1}{2}h'\left(\frac{u-v}{2}\right)=\frac{f'(u)}{f(u)}$$

上式两边再关于 v 求偏导得

$$\frac{1}{4}h''\left(\frac{u+v}{2}\right)-\frac{1}{4}h''\left(\frac{u-v}{2}\right)=0$$

令 $v=u$ 得

$$h''(u)-h''(0)=0$$

记 $h''(0)=-\lambda$, 又改写 u 为 x, 则得

$$\ln p(x)=-\frac{\lambda}{2}x^2+C_1x+C_2$$

于是

$$p(x)=\mathrm{e}^{-\frac{\lambda}{2}x^2+C_1x+C_2}=\mathrm{e}^{C_2+\frac{C_1^2}{2\lambda}}\,\mathrm{e}^{-\frac{\lambda}{2}(x-\frac{C_1}{\lambda})^2}=K\,\mathrm{e}^{-\frac{\lambda}{2}(x-\frac{C_1}{\lambda})^2}$$

因 $p(x)$ 是密度函数, 故有 $\lambda>0$. 此外, $C_1=\dfrac{p'(0)}{p(0)}$, $C_2=\ln p(0)$.

由此可见 ξ,η 服从正态分布, 从而 $\xi+\eta,\xi-\eta$ 也服从正态分布.

【评注】本题含意深刻, 指出正态分布的一种重要刻画.

44. 把习题 25 和习题 26 的结果推广到 n 个随机变量的场合.

答 (1) 若 $\xi_1,\cdots,\xi_n$ 是 n 个相互独立随机变量, 且 $\xi_i\sim B(m_i,p)$, 则

$$\xi_1+\cdots+\xi_n\sim B(m_1+\cdots+m_n,p)$$

$$P\{\xi_1=\ell_1,\cdots,\xi_n=\ell_n\,|\,\xi_1+\cdots+\xi_n=\ell\}=\frac{\dbinom{m_1}{\ell_1}\cdots\dbinom{m_n}{\ell_n}}{\dbinom{m_1+\cdots+m_n}{\ell}}$$

其中, 整数 $\ell_i\geqslant 0$, $i=1,2,\cdots,n$, $\ell_1+\cdots+\ell_n=\ell$.

(2) 若 $\xi_1,\cdots,\xi_n$ 是相互独立随机变量, 均服从泊松分布, 参数

分别为 $\lambda_1,\cdots,\lambda_n$, 则

$$\xi_1+\cdots+\xi_n\sim P(\lambda_1+\cdots+\lambda_n)$$

$$P\{\xi_1=k_1,\cdots,\xi_n=k_n\,|\,\xi_1+\cdots+\xi_n=k\}$$
$$=\frac{k!}{k_1!\cdots k_n!}\Big(\frac{\lambda_1}{\lambda_1+\cdots+\lambda_n}\Big)^{k_1}\cdots\Big(\frac{\lambda_n}{\lambda_1+\cdots+\lambda_n}\Big)^{k_n}$$

其中整数 $k_i\geqslant 0$, $i=1,2,\cdots,n$, $k_1+\cdots+k_n=k$.

【评注】二项分布与泊松分布的再生性是可预期的结果. 有趣的是条件分布: 在二项分布场合得到多元超几何分布; 在泊松场合则得到多项分布.

45. 设随机变量 ξ 与 η 相互独立, $\xi\sim B(n,p)$, 而 η 服从 $(0,1)$ 上均匀分布, 试求 $\xi+\eta$ 的分布函数和密度函数.

解 记

$$b_l=\binom{n}{l}p^lq^{n-l},\quad l=0,1,2,\cdots,n$$

对 $x\leqslant 0$,

$$P\{\xi+\eta<x\}=0$$

对 $x>n+1$,

$$P\{\xi+\eta<x\}=1$$

对 $k<x\leqslant k+1$, $k=0,1,2,\cdots,n$,

$$\begin{aligned}P\{\xi+\eta<x\}&=P\{\xi+\eta<x,\ \xi\leqslant k-1\}+P\{\xi+\eta<x,\ \xi=k\}\\&\quad+P\{\xi+\eta<x,\ \xi\geqslant k+1\}\\&=P\{\xi\leqslant k-1\}+P\{\xi+\eta<x,\ \xi=k\}\\&=\sum_{l=0}^{k-1}b_l+P\{\xi=k\}P\{\xi+\eta<x\,|\,\xi=k\}\\&=\sum_{l=0}^{k-1}b_l+b_kP\{\eta<x-k\,|\,\xi=k\}\end{aligned}$$

$$= \sum_{l=0}^{k-1} b_l + b_k P\{\eta < x - k\}$$

$$= \sum_{l=0}^{k-1} b_l + b_k (x - k)$$

即分布函数

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ \sum_{l=0}^{k-1} b_l + b_k(x-k), & k < x \leqslant k+1, \quad k = 0, 1, \cdots, n \\ 1, & x > n+1 \end{cases}$$

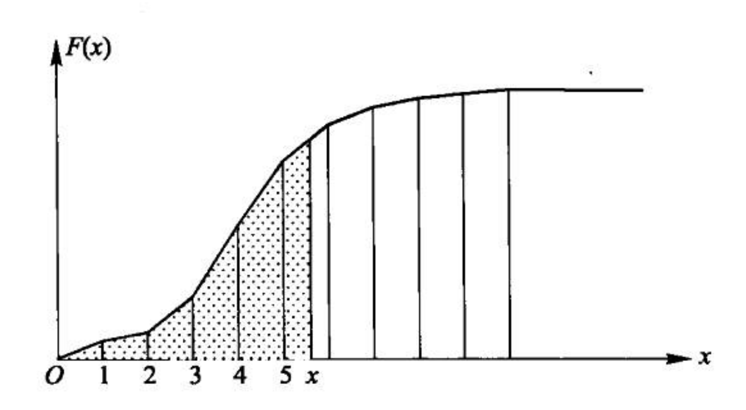


图 3-6

密度函数为

$$p(x) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k < x \leqslant k+1, \quad k = 0, 1, 2, \cdots, n \\ 0, & x \leqslant 0, \quad x > n+1 \end{cases}$$

【评注】讨论离散型变量与连续型变量之和, 别具一格.

以上论证过程显然可推广到任意离散型变量与连续型变量之和.

**46. 试求顺序统计量 ξ_k^* 与 ξ_l^* $(k < l)$ 的联合密度函数.

解 对 $x<y$, 见图 3-7

$P\{x \leqslant \xi_k^* < x+\Delta x,\ y \leqslant \xi_l^* < y+\Delta y\}$ (化为多项分布)

$$= \frac{n!}{(k-1)!\,1!\,(l-k-1)!\,1!\,(n-l)!}\big[F(x)\big]^{k-1}p(x)\Delta x$$

$$\cdot\big[F(y)-F(x+\Delta x)\big]^{l-k-1}p(y)\Delta y\big[1-F(y+\Delta y)\big]^{n-l}$$

图 3-7

故

$$p_{(\xi_k^*,\xi_l^*)}(x,y)=\begin{cases}\dfrac{n!}{(k-1)!(l-k-1)!(n-l)!}\big[F(x)\big]^{k-1} \\ \quad\cdot\big[F(y)-F(x)\big]^{l-k-1}\big[1-F(y)\big]^{n-l}p(x)p(y), & x<y \\ 0, & x\geqslant y\end{cases}$$

【评注】但愿读者学会这种既直观又便于记忆的方法. 50 年笔者于 51 年前从匈牙利专家 Vincze István 处学到.

47. 试利用概率的连续性重新证明一元分布函数的性质 (ii) 和性质 (iii), 并说明这种证法可推广到多元的场合.

证 关于性质(ii):

记 $A_m=\{\xi<-m\}$, $m=1,2,3,\cdots$, 则

$$A_m\supset A_{m+1},\quad \lim_{m\to\infty}A_m=\bigcap_{m=1}^{\infty}A_m=\varnothing$$

而根据概率的上连续性知

$$\lim_{m\to\infty}F(-m)=\lim_{m\to\infty}P\{\xi<-m\}=\lim_{m\to\infty}P(A_m)=P(\lim_{m\to\infty}A_m)$$

$$=P\Big(\bigcap_{m=1}^{\infty}A_m\Big)=0$$

利用 $F(x)$ 的单调性即知 $\lim\limits_{x\to-\infty}F(x)=0$.

同理可证 $\lim\limits_{x\to+\infty}F(x)=1$.

关于性质 (iii):

记 $A_m=\left\{x_0-\dfrac{1}{m}\leqslant\xi<x_0\right\}$, $m=1,2,3,\cdots$, 则

$$A_m\supset A_{m+1},\quad \lim_{m\to\infty}A_m=\bigcap_{m=1}^{\infty}A_m=\varnothing$$

于是

$$\lim_{m\to\infty}\left[F(x_0)-F\left(x_0-\frac{1}{m}\right)\right]=\lim_{m\to\infty}P(A_m)=P(\lim_{m\to\infty}A_m)$$
$$=P(\varnothing)=0$$

即

$$F(x_0)=\lim_{m\to\infty}F\left(x_0-\frac{1}{m}\right)$$

仍由 $F(x)$ 的单调性知分布函数左连续.

对 n 维随机向量 $(\xi_1,\cdots,\xi_n)$ 的分布函数 $F(x_1,\cdots,x_n)$ 只证 $F(+\infty,\cdots,+\infty)=1$.

记 $A_m=\{\xi_1<k_1+m,\cdots,\xi_n<k_n+m\}$, $m=1,2,\cdots$, 则

$$A_m\subset A_{m+1},\quad \lim_{m\to\infty}A_m=\bigcup_{m=1}^{\infty}A_m=\Omega$$

根据概率的下连续性知

$$\begin{aligned}&\lim_{m\to\infty}F(k_1+m,\cdots,k_n+m)\\&\qquad=\lim_{m\to\infty}P\{\xi_1<k_1+m,\cdots,\xi_n<k_n+m\}\\&\qquad=\lim_{m\to\infty}P(A_m)=P(\lim_{m\to\infty}A_m)=P(\Omega)=1\end{aligned}$$

利用分布函数的单调性知

$$\lim_{\substack{x_1\to+\infty\\ \cdots\cdots\\ x_n\to+\infty}}F(x_1,\cdots,x_n)=1$$

其余部分请读者补足.

【评注】《概率论基础》用可列可加性证明 (ii), (iii), 好处是只用概率三公理, 而且可看出分布函数三性质与概率三公理的对应, 但在多元场合, 用作为公理推论的概率的连续性似乎更方便.

*48. 利用随机变量分布解释贝特朗奇论.

解 参看《概率论基础》图 1.4.4.

(1) 将弦的一端 A 固定, 另一端 B 在圆周上选取. 以 ξ_1 表示沿逆时针方向 $\widehat{AB}$ 弧长, 若 ξ_1 在 $(0,2\pi)$ 上服从均匀分布, 则

$$P\{\text{弧长} > \sqrt{3}\} = P\left\{\frac{2\pi}{3} < \xi_1 < \frac{4\pi}{3}\right\} = \frac{1}{3}$$

(2) 假定弦垂直于某直径, 取该直径为 y 轴, 圆心为坐标原点, 以 ξ_2 表示弦的中点坐标, 若 ξ_2 在 $[-1,1]$ 上服从均匀分布, 则

$$P\{\text{弦长} > \sqrt{3}\} = P\left\{-\frac{1}{2} < \xi_2 < \frac{1}{2}\right\} = \frac{1}{2}$$

(3) 以圆心为原点建立直角坐标系 XOY, 记弦的中点坐标为 $\eta = (\eta_1, \eta_2)$, 若 η 在圆内 $\{(x,y) | x^2+y^2 \leqslant 1\}$ 服从均匀分布, 记 $D = \left\{(x,y) \middle| x^2+y^2 < \frac{1}{2}\right\}$, 则

$$P\{\text{弦长} < \sqrt{3}\} = P\{\eta \in D\} = \frac{1}{4}$$

三种解法中的随机变量服从三种不同的均匀分布假设, 所以得出不同的结论.

【评注】至此总算对贝特朗奇论有个交待.

*49. 若 f 是 Ω 上单值实函数, 对 $B \subset \mathbf{R}^1$, 记 $f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$, 试证逆映照 f^{-1} 具有如下性质:

(1) $f^{-1}\left(\bigcup_{\lambda\in\Lambda} B_\lambda\right) = \bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$;

(2) $f^{-1}\left(\bigcap_{\lambda\in\Lambda} B_\lambda\right) = \bigcap_{\lambda\in\Lambda} f^{-1}(B_\lambda)$;

(3) $f^{-1}(\overline{B}) = \overline{f^{-1}(B)}$.

证 (1) 若 $\omega \in f^{-1}\left(\bigcup_{\lambda\in\Lambda} B_\lambda\right)$, 则 $f(\omega) \in \bigcup_{\lambda\in\Lambda} B_\lambda$, 必存在 $\lambda_0 \in \Lambda$ 使 $f(\omega) \in B_{\lambda_0}$, 亦有 $\omega \in f^{-1}(B_{\lambda_0})$, 从而 $\omega \in \bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$.

所以, $f^{-1}\left(\bigcup_{\lambda\in\Lambda} B_\lambda\right) \subset \bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$.

反之, 若 $\omega \in \bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$, 必存在 $\lambda_0 \in \Lambda$ 使 $\omega \in f^{-1}(B_{\lambda_0})$, 亦有 $f(\omega) \in B_{\lambda_0}$, 即 $f(\omega) \in \bigcup_{\lambda\in\Lambda} B_\lambda$, 从而 $\omega \in f^{-1}\left(\bigcup_{\lambda\in\Lambda} B_\lambda\right)$.

$\therefore f^{-1}\left(\bigcup_{\lambda\in\Lambda} B_\lambda\right) \supset \bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$.

于是有 $f^{-1}\left(\bigcup_{\lambda\in\Lambda} B_\lambda\right) = \bigcup_{\lambda\in\Lambda} f^{-1}(B_\lambda)$.

(2), (3) 可类似证明. 留给读者作为练习.

【评注】 为证两个事件等价采用由定义出发的证明, 看似笨拙, 但舍此并无他法.

**50. 证明: ξ 是一个随机变量当且仅当对任何 $x \in \mathbf{R}^1$, 成立

$$\{\omega : \xi(\omega) < x\} \in \mathscr{F}$$

[提示: 必要性是显然的, 为证充分性, 记 $\mathfrak{m} = \{A : A \subset \mathbf{R}^1, (\xi(\omega) \in A) \in \mathscr{F}\}$, 验证 $\mathfrak{m}$ 是 σ 域, 又 $\mathfrak{m}$ 包含全体形如 $(-\infty, x)$ 的区间, 故 $\mathfrak{m}$ 包含 $\mathscr{B}_1$]

证 必要性: 设 ξ 是随机变量, 则按《概率论基础》定义 (3.1.1) 对任一 $B \in \mathscr{B}_1$ 有 $\{\omega : \xi(\omega) \in B\} \in \mathscr{F}$, 又 $(-\infty, x) \in \mathscr{B}_1$, 有

$$\{\omega : \xi(\omega) < x\} = \{\omega : \xi(\omega) \in (-\infty, x)\} \in \mathscr{F}$$

充分性: 记 $\mathfrak{m} = \{A : A \subset \mathbf{R}^1, (\omega : \xi(\omega) \in A) \in \mathscr{F}\}$, 现证 $\mathfrak{m}$ 是 $\mathbf{R}^1$上的 σ 域.

(1) $\{\omega : \xi(\omega) \in \mathbf{R}^1\} = \Omega \in \mathscr{F}$, 故 $\mathbf{R}^1 \in \mathfrak{m}$;

(2) 若 $B \in \mathbf{m}$, 由 49 题 (3) $f^{-1}(\overline{B}) = \overline{f^{-1}(B)}$ 得

$$\{\omega : \xi(\omega) \in \overline{B}\} = \Omega - \{\omega : \xi(\omega) \in B\} \in \mathscr{F}$$

即 $\overline{B} \in \mathbf{m}$, 故 $\mathbf{m}$ 对余集运算封闭;

(3) 设 $B_i \in \mathbf{m}$, $i = 1, 2, \cdots$, 由 49 题 (1) 的结论得

$$\{\omega :\ \xi(\omega) \in \bigcup_{i=1}^{\infty} B_i\} = \bigcup_{i=1}^{\infty} \{\omega :\ \xi(\omega) \in B_i\} \in \mathscr{F}$$

即 $\bigcup\limits_{i=1}^{\infty} B_i \in \mathbf{m}$, 故 $\mathbf{m}$ 关于可列并集运算封闭.

由 (1)—(3) 知, $\mathbf{m}$ 是 σ 域, 由条件知

$$\{(-\infty, x) : x \in \mathbf{R}^1\} \subset \mathbf{m}$$

所以, $\mathbf{m} \supset \sigma\{(-\infty, x) : x \in \mathbf{R}^1\} = \mathscr{B}_1$

其中 $\sigma\{A\}$ 表示由集类 A 产生的 σ 域, 由此得证 ξ 是一随机变量.

【评注】本题完成了《概率论基础》中引用未证的第一个测度论结论 (见原书 119 - 120 页) 的证明, 用的是典型的测度论方法, 很新颖但并不难.

显然, 这个证明只要稍作符号替换就能用于随机向量场合, 这就证得该书第二个引用待证的测度论结论 (见原书 142 - 143 页).

《概率论基础》不把证明直接写入正文而编入本章最后两道习题, 自有深层考虑.

习题总评

第三章习题的主角是分布. 大部分习题围绕分布的表示, 分布的计算, 分布的性质, 分布的推导而配置.

前两章的习题主要是计算, 求正确的答案. 本章则主要是演算, 用微积分及级数为工具作多种分析推导. 不少题还是以证明题的形式出现, 题中已指明了答案. 总的说, 做出结果并不太难, 但每个题目背后的意义则常为学生们所忽视, 因此希望同学们在完成每道

题之后务必要花一定时间思考一下题目的含意和回顾一下解题的思路, 才不会 "空入宝山".

章后小议

随机变量的引进使对随机现象的处理更简单更直接, 也更统一而有力, 是观念上的一大进步, 随机变量概念的引进也使概率论融入以分析为主流的数学, 命题表述变得清晰, 研究工具大量增加, 讲授进度显著增快.

在公理化体系中, 随机变量定义为样本空间到欧几里得空间的可测映照. 这样一来, 事件用随机变量的某个关系式表示, 概率则通过分布函数计算, 许多概率问题化为分析问题.

当然上述定义有很深奥的测度论背景, 但单从操作层面上看, 对随机变量就可以进行加减乘除等算术运算和微分, 积分及求极限等分析运算, 而分布函数则是分析性质良好的单调函数.

分布函数成了研究的中心. 很多研究成果通过各个分布函数积累起来. 理清这些材料并使之系统化, 在本章正文和习题中都有所体现, 尤其让离散型与连续型等待时间分布以对称的顺序出场, 在《概率论基础》之前的概率论教本中似未见到.

在多维场合, 随机向量由多元联合分布完整描述, 而边际分布, 条件分布等新概念的引进更是丰富了研究主题. 不过, 有名的多元分布比起一元分布少很多, 除了出于摸球模型的多项分布和多元超几何分布外, 多元正态与联系几何概率的均匀分布是主要的例子.

独立性概念从事件推广到随机变量, 并在许多讨论中起关键作用. 列联表对比了两类摸球模型, 而矩形均匀分布则回归到几何概率.

概率论中的主角正态分布终于在本章登场, 受到最大的重视, 给出了高斯推导, 还有典型分解, 逐步走向舞台中央.

随机变量的函数仍然是随机变量有了明确的含义, 因而又回到

求分布函数的老问题上来, 这是本章技术训练的重点.《概率论基础》采用一对一, 多对一, 多对多三个类型与直接法, 变换法两种方法交叉的形式, 既对方法作了比较全面的介绍, 又导出不少重要的结果, 特别是给出统计学三大分布: χ^2, t, F 的理论推导.

均匀分布作为分布函数转换器的特殊地位被揭示. 顺带证明了随机变量的存在性定理, 这类定理一般被人们在不知不觉中使用, 很少点明. 从数学的形式逻辑观点看, 具有某分布函数的随机变量的存在性问题不能说不重要.

极小值与极大值联合分布的推导为统计学作了重要准备.

教学札记之十一

柯西方程与分布刻画

《概率论基础》引理 2.4.1 证明函数方程 $f(x)f(y) = f(x+y)$ 若对一切 x, y (或一切 $x \geqslant 0, y \geqslant 0$) 成立, 则只要假定 $f(x)$ 是连续函数或单调函数, 其解必为 $f(x) = a^x$, 其中 $a \geqslant 0$, 为常数.

事实上这类方程共有 4 个,

$$f(x) + f(y) = f(x+y) \tag{1}$$

$$f(x) \cdot f(y) = f(x+y) \tag{2}$$

$$f(x) + f(y) = f(xy) \tag{3}$$

$$f(x) \cdot f(y) = f(xy) \tag{4}$$

它们通称为柯西方程. 可以证明: 若它们在一定范围成立, 且 $f(x)$ 满足微弱的条件 (例如只要假定满足 (1) 的函数 $f(x)$ 是可测函数), 则这 4 个函数方程的解分别是

$$f(x) = ax \tag*{(1)$'$}$$

$$f(x) = b^x, \quad b \geqslant 0 \tag*{(2)$'$}$$

$$f(x) = \log_c x, \quad c > 0 \tag*{(3)$'$}$$

$$f(x) = x^\alpha \tag*{(4)$'$}$$

正好对应 4 种初等函数. 方程 (1) (2) 可以互化, 因此实质上只要研究一个方程,《概率论基础》择便选了方程 (2).

显然上述某些结果可推广到 x, y 只取整数的场合. 例如可参看《概率论基础》引理 4.3.1.

分布刻画是概率论的一个研究专题. 人们发现, 只要具有某个性质就一定是某某分布, 例如若具有无记忆性, 则连续型非负随机变量一定服从指数分布, 离散型则对应为几何分布. 因此寻找各种分布具有的特征性质便成为一个研究热门, 就产生了分布刻画这一研究专题. 这个研究专题, 上承概率论, 下接数理统计, 具有重要的理论意义和实用价值.

在分布刻画的研究中常归结到柯西方程.《概率论基础》重视对分布个性的介绍, 因此多处涉及这些内容, 所以引入引理 2.4.1.

相关内容有:《概率论基础》中第二章泊松过程推导中 $P_0(t)$ 的导出; 第三章 §1 几何分布与指数分布的刻画; §2 正态分布的高斯推导; 习题三 10 题均匀分布的刻画; 习题三 **43 题对正态分布的刻画. 有关的尚有第四章熵的表达式的推导以及用最大熵刻画均匀分布、指数分布及正态分布, 还有习题四 **43 题对泊松分布的刻画.

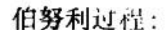
教学札记之十二

关于两个等待时间分布序列

一、两个等待时间序列

若每隔 Δt 进行一次试验, 则伯努利试验可以看作一个随时间而变化的随机过程, 关心的是到时刻 t 成功出现的次数, 记作 $X(t)$, $t=\Delta t, 2\Delta t, \cdots, n\Delta t$, 称为成功概率为 p 的伯努利过程.
以 $Y(t), t\geqslant 0$ 记参数为 λt 的泊松过程到时刻 t 来到的呼叫数.

$X(t)$ 与 $Y(t)$ 随 t 变化的图形都是跳跃高度为 1 的上升阶梯函数, 其跳跃时刻分别为成功出现与呼叫到达的时刻.



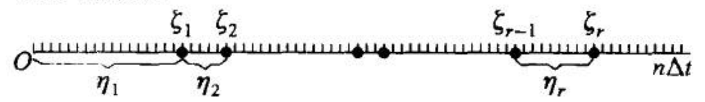


图 3－8

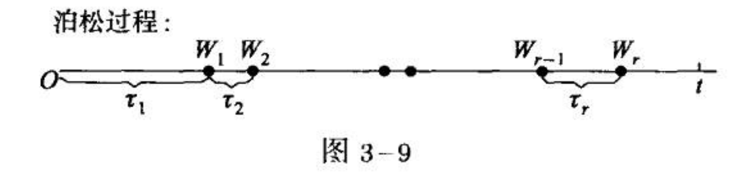


图 3－9

从《概率论基础》第二章的讨论可知,

$$P\{X(n\Delta t)=k\}=\binom{n}{k}p^kq^{n-k},\quad k=0,1,2,\cdots,n \tag{1}$$

而

$$P\{Y(t)=k\}=\frac{(\lambda t)^k}{k!}\,\mathrm{e}^{-\lambda t},\quad k=0,1,2,\cdots \tag{1$'$}$$

从第二章 §3 及第三章 §1 的讨论又知道:

若以 η_1 记伯努利过程第一次 (首次) 成功出现在 $k\Delta t$, 其概率服从几何分布

$$q^{k-1}p,\quad k=1,2,\cdots \tag{2}$$

若以 τ_1 记泊松过程第一个 (首个) 呼叫出现时刻, 其概率分布为指数分布

$$\lambda\,\mathrm{e}^{-\lambda t},\quad t\geqslant 0 \tag{2$'$}$$

若以 ζ_r 记伯努利过程第 r 次成功出现在 $k\Delta t$, 其概率服从帕斯卡分布

$$\binom{k-1}{r-1}p^rq^{k-r},\quad k=r,r+1,\cdots \tag{3}$$

若以 W_r 记泊松过程第 r 个呼叫出现时刻, 其概率分布为埃尔朗分布

$$\frac{\lambda(\lambda t)^{r-1}}{(r-1)!}\,\mathrm{e}^{-\lambda t},\quad t\geqslant 0 \tag{3$'$}$$

帕斯卡分布与埃尔朗分布分别可以推广成对任意实数 $r>0$ 都有意义的负二项分布及 Γ 分布.

二、两个分解式

在伯努利过程中, 第 r 个到达时刻 ζ_r 有如下分解式:

$$\zeta_r=\eta_1+\eta_2+\cdots+\eta_r \tag{4}$$

其中 η_i 表示从第 $i-1$ 次成功到第 i 次成功之间的试验次数, 简称第 i 次等待时间. 其中第 1 次等待时间 η_1 前已定义, 并知道它服从几何分布 (2). 下面我们要进一步证明:

定理 1 $\eta_1,\eta_2,\cdots,\eta_r$ 相互独立, 均服从几何分布 (2).

证明 由 (4) 知

$$\begin{aligned}&\eta_1=\zeta_1\\&\eta_2=\zeta_2-\zeta_1\\&\cdots\cdots\cdots\cdots\\&\eta_r=\zeta_r-\zeta_{r-1}\end{aligned} \tag{5}$$

对任意正整数 $l_1<l_2$,

$$\begin{aligned}&P\{\eta_1=l_1,\eta_2=l_2-l_1\}=P\{\zeta_1=l_1,\zeta_2=l_2\}\\&=P\big\{X(\Delta t)=0,\cdots,X\big((l_1-1)\Delta t\big)=0,X(l_1\Delta t)=1,\\&\qquad\cdots,X\big((l_2-1)\Delta t\big)=1,X(l_2\Delta t)=2\big\}\\&=\underbrace{q\cdots q}_{l_1-1}\,p\,\underbrace{q\cdots q}_{l_2-l_1-1}\,p\\&=q^{l_1-1}\cdot p\cdot q^{l_2-l_1-1}\cdot p\\&=P\{\eta_1=l_1\}P\{\eta_2=l_2-l_1\}\end{aligned}$$

一般地, 对正整数 $l_1<l_2<\cdots<l_r$,

$$\begin{aligned}&P\{\eta_1=l_1,\eta_2=l_2-l_1,\cdots,\eta_r=l_r-l_{r-1}\}\\&=P\{\zeta_1=l_1,\zeta_2=l_2,\cdots,\zeta_r=l_r\}\\&=q^{l_1-1}p\,q^{l_2-l_1-1}p\cdots q^{l_r-l_{r-1}-1}p\\&=P\{\eta_1=l_1\}P\{\eta_2=l_2-l_1\}\cdots P\{\eta_r=l_r-l_{r-1}\}\end{aligned} \tag{6}$$

由 $l_1, l_2, \cdots, l_r$ 的任意性知 $\eta_1, \eta_2, \cdots, \eta_r$ 相互独立, 均服从参数为 p 的几何分布.

同样地, 对泊松过程, 第 r 个到达时刻 W_r 有如下分解式:

$$W_r = \tau_1 + \tau_2 + \cdots + \tau_r \tag{4$'$}$$

其中 τ_i 表示过程从第 $i-1$ 次到达到第 i 次到达之间的时间间隔, 也简称第 i 次等待时间, 其中第 1 次等待时间 τ_1 前已定义, 并知道它服从指数分布 (2)$'$. 下面进一步证明:

定理 1$'$ $\tau_1, \tau_2, \cdots, \tau_r$ 相互独立, 均服从指数分布 (2)$'$.

证明 记

$$\begin{aligned} \tau_1 &= W_1 \\ \tau_2 &= W_2 - W_1 \\ &\cdots\cdots\cdots\cdots \\ \tau_n &= W_n - W_{n-1} \end{aligned} \tag{7}$$

下面来导出 $(\tau_1, \tau_2, \cdots, \tau_n)$ 的联合密度函数.

我们先来导出 $(W_1, W_2, \cdots, W_n)$ 的联合密度函数, 下面推荐一个简捷的做法. 选

$$s_1 < s_1 + \Delta s_1 < s_2 < s_2 + \Delta s_2 < \cdots < s_n < s_n + \Delta s_n$$

图 3-10

则利用泊松过程的独立增量性、平稳性和关于增量分布为泊松分布的结论, 可知

$$\begin{aligned} & P\{s_1 \leqslant W_1 < s_1 + \Delta s_1, s_2 \leqslant W_2 < s_2 + \Delta s_2, \cdots, s_n \leqslant W_n < s_n + \Delta s_n\} \\ &= \mathrm{e}^{-\lambda s_1} \times \frac{\lambda(s_1 + \Delta s_1 - s_1)}{1!} \mathrm{e}^{-\lambda(s_1 + \Delta s_1 - s_1)} \\ &\quad \times \mathrm{e}^{-\lambda(s_2 - s_1 - \Delta s_1)} \times \frac{\lambda \Delta s_2}{1!} \mathrm{e}^{-\lambda \Delta s_2} \end{aligned} \tag{8}$$

$$\times \cdots \times e^{-\lambda(s_n - s_{n-1} - \Delta s_{n-1})} \times \frac{\lambda \Delta s_n}{1!} e^{-\lambda \Delta s_n}$$

$$= \lambda^n e^{-\lambda s_n} \Delta s_1 \Delta s_2 \cdots \Delta s_n + o(\Delta s_1 \Delta s_2 \cdots \Delta s_n)$$

因此

$$p_{(W_1, W_2, \cdots, W_n)}(s_1, s_2, \cdots, s_n) = \begin{cases} \lambda^n e^{-\lambda s_n}, & s_1 < s_2 < \cdots < s_n \\ 0, & \text{其他} \end{cases} \tag{9}$$

用变换法得到

$$\begin{aligned} q_{(\tau_1, \tau_2, \cdots, \tau_n)}(t_1, t_2, \cdots, t_n) &= \lambda^n e^{-\lambda(t_1 + t_2 + \cdots + t_n)} \\ &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda t_2} \cdots \lambda e^{-\lambda t_n} \end{aligned} \tag{10}$$

故知 $\tau_1, \tau_2, \cdots, \tau_n$ 相互独立, 均服从 Exp (λ), 从而完成了证明.

当然上述过程也容易用来证明逆命题, 即若 $\tau_1, \tau_2, \cdots, \tau_n$ 相互独立, 均服从 Exp (λ), 则 $(W_1, W_2, \cdots, W_n)$ 服从 (9), 不过这个逆命题还有多种多样的证法.

顺便指出, 由 (9) 也能导出 W_r 的边际分布密度

$$p_{W_r}(s_r) = \frac{\lambda^r s_r^{r-1}}{(r-1)!} e^{-\lambda s_r}, \quad r = 1, 2, \cdots$$

至此我们已对《概率论基础》作了适当的补充.

三、两列分布对比

两列等待时间分布有许多共性, 列表对比于下:

分布	概率分布/密度函数	均值	方差	特征函数	特征刻画
几何分布	$q^{k-1}p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{p e^{it}}{1 - q e^{it}}$	无记忆
指数分布	$\lambda e^{-\lambda x}$	λ^{-1}	λ^{-2}	$\left(1 - \frac{it}{\lambda}\right)^{-1}$	无记忆
帕斯卡分布	$\binom{k-1}{r-1} p^r q^{k-r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$\left(\frac{p e^{it}}{1 - q e^{it}}\right)^r$	分解式 (4)
埃尔朗分布	$\frac{\lambda^r x^{r-1}}{(k-1)!} e^{-\lambda x}$	$r\lambda^{-1}$	$r\lambda^{-2}$	$\left(1 - \frac{it}{\lambda}\right)^{-r}$	分解式 (4)′

四、附言

《概率论基础》第二章§3 给出了伯努利试验的描述, §4 给出了泊松过程的结构分析.

定理 2.4.1 给出泊松逼近定理及其直接证明, 第四章§4 例 7 用母函数法重加证明.

第二章§3 给出几何分布、帕斯卡分布的定义.

第三章§1 讨论了几何分布、帕斯卡分布、指数分布、埃尔朗分布的定义与性质, 并推广到负二项分布及 Γ 分布.

习题三 *8 题要求学生证明用埃尔朗分布逼近帕斯卡分布.

第五章§2 用特征函数法再次研究上述逼近.

教学札记之十三

摸球与抽样

《概率论基础》中摸球模型多次出现, 许多重要的概念是经它引入的.

摸球模型作为抽样的模型, 首先出现在古典概型中. 在装有 a 只黑球 b 只白球的袋中摸出一球, 求摸得黑球的概率是典型的古典概率计算. 接着自然地考虑多次模球的情形, 这时发现必须区分放回与不放回两种方式. 首先证明了 "抽签与顺序无关", 即第 k 次摸得黑球的概率为常数, 与 k 无关, 这是对不放回摸球讲的, 因为对放回摸球这是不言而喻的.

其次, 导出了二项分布与超几何分布这两个重要的分布, 前者针对放回摸球, 后者针对不放回摸球, 显示了二者的重大区别.

第二章考察了摸球问题的前后结果事件 A 与 B, 发现其条件概率在放回摸球与不放回摸球两种情况下有重要差别, 这启发引入独立性概念.

n 次有放回摸球是伯努利试验的重要特例, 在这个试验中二项分布成为研究中心, 还能导出其他分布, 例如几何分布与帕斯卡分

布.

第三章用随机变量 (ξ,η) 描述摸球问题, 讲了列联表——表 3.2.1 与 3.2.2, 它形象地告诉我们: 放回摸球的各次结果是独立同分布的, 而不放回摸球是不独立的, 但还是同分布的.

把双色球推广到多色球, 这时放回摸球模型成了推广的伯努利试验, 导出多项分布, 而不放回摸球则导出多元超几何分布.

若每个球对应一个数, 则摸球问题对应抽样调查这一社会科学中最主要的统计方法. 所需计算, 在放回场合各变量独立, 容易处理; 而在不放回抽样这一实用场合, 各变量不独立, 它们的相关性使计算复杂化, 但《概率论基础》第四章§2 例 10 介绍了这种场合的巧妙处理方法, 并顺带导出超几何分布的均值与方差表达式.

对于随机变量 X 取无穷个值的情况, 这时不宜再用摸球模型的形象与术语, 但抽样的术语照用, 放回与不放回已无差别, 因此统计学中把独立同分布观察样本 (称为简单随机样本, 可以来自有限总体或无限总体) 作为研究对象建立起当代数理统计学, 而把只研究有限总体的抽样调查作为一个重要分支.

总之, 摸球作为有限总体抽样的形象模型, 概要如下表:

摸球(抽样)	放回	独立	同分布	⟶	数理统计
	不放回	不独立	同分布	⟶	抽样调查

摸球模型还可以有许多变形, 例如在第二章波利亚坛子模型中, 通过中间加入若干个某色球来改变摸出各色球的概率, 成为很有用的新模型, 该模型不但被用于传染病学研究, 还成为一种表达力很强的随机过程.

教学札记之十四

正态分布的两种刻画

自高斯以样本均值为总体均值的最大似然估计导出正态分布

以来, 逐渐发现正态分布的另外一些刻画性特征, 这里介绍其中最重要的两种.

射击弹落点分布 在火器射击问题中, 假如过射击目标中心作直角坐标系的 x 轴及 y 轴. 每次射击弹落点由于受随机因素影响会偏离目标, 它的坐标 (ξ,η) 是一个二维随机变量, 假定满足如下三个条件:

(1) ξ 与 η 分别具有连续密度函数 $p_1(x)$ 及 $p_2(y)$;

(2) ξ 与 η 相互独立;

(3) (ξ,η) 的密度函数在 (x,y) 点的值仅与它到原点的距离 $r=\sqrt{x^2+y^2}$ 有关,

则 ξ 与 η 均服从正态分布.

下面我们就来证明这个结论. 为不打断证明思路, 特把用到的一个分析结果写成引理, 它本身也颇有意思.

引理 若 $f(x)$ 及 $g(y)$ $(x\geqslant 0,\ y\geqslant 0)$ 是不恒等于 0 的连续函数, 而且对一切非负实数 x 及 y 成立

$$f(x)g(y)=h(x+y) \tag{$*$}$$

则

$$f(x)=ka^x,\quad x\geqslant 0 \tag{$**$}$$

这里 k 及 a 是常数, 而且 $a>0$.

证明 首先证明 $f(0)\neq 0$. 用反证法, 假定 $f(0)=0$, 则 $h(y)\equiv 0$, 由于 $g(y)$ 不恒为 0, 有一个 y_0 使 $g(y_0)\neq 0$, 因此 $f(x)=\dfrac{h(x+y_0)}{g(y_0)}\equiv 0$, 这与 $f(x)$ 不恒等于 0 的假设矛盾, 从而证得 $f(0)\neq 0$, 同理可证 $g(0)\neq 0$.

其次, 在 $(*)$ 中令 $x=z$, $y=0$, 得到 $f(z)g(0)=h(z)$, 又在 $(*)$ 中令 $x=0$, $y=z$, 得到 $f(0)g(z)=h(z)$, 因此

$$f(z)g(0)=f(0)g(z)=h(z)$$

两边除以 $f(0)g(0)$ 并记得到的比值为 $p(z)$, 则得

$$\frac{f(z)}{f(0)}=\frac{g(z)}{g(0)}=\frac{h(z)}{f(0)g(0)}=p(z)$$

代入 (*) 式得

$$p(x)p(y)=p(x+y)$$

利用《概率论基础》的引理 2.4.1 知

$$p(x)=a^x,\quad a>0$$

因此

$$f(x)=f(0)a^x$$

引理得证.

下面来证明主要结果.

记 (ξ,η) 的密度函数为 $p(x,y)$, 由假定 (3) 可知 $p(x,y)=q\left(\sqrt{x^2+y^2}\right)$, 再由假定 (1), (2) 知

$$p_1(x)p_2(y)=q\left(\sqrt{x^2+y^2}\right)$$

因为 $p_1(x)$ 及 $p_2(y)$ 为密度函数, 当然不能恒等于 0, 故可利用引理 (把 $p_1(x)$ 看作 $f(x^2)$, 把 $p_2(y)$ 看作 $g(y^2)$, 再把 $q(\sqrt{x^2+y^2})$ 看作 $h(x^2+y^2)$), 得到

$$p_1(x)=ka^{x^2}$$

由于 $p_1(x)$ 是 $(-\infty,\infty)$ 上的密度函数, 故有 $0<a<1$, 记 $a=\mathrm{e}^{-\frac{1}{2\sigma_1^2}}$, 其中 $\sigma_1>0$, 利用规范化条件可得 $k=\dfrac{1}{\sqrt{2\pi}\sigma_1}$, 这样一来,

$$p_1(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{x^2}{2\sigma_1^2}}$$

因此 ξ 服从正态分布, 同理可知 η 也服从正态分布.

以上内容见于《概率论基础》第一版及第二版. 在第三版中它为高斯推导所代替.

气体分子速度分布 1859 年麦克斯韦在气体分子运动的研究中, 在很简单的假定下也导出正态分布.

若以 C 代表分子速度, 它在 x, y, z 方向的三个分量分别记为 U, V, W, 则 $C^2 = U^2 + V^2 + W^2$. 麦克斯韦对这些随机变量的分布提出如下三个假定:

(1) 速度分布存在具有连续导函数的密度函数;

(2) U, V, W 相互独立;

(3) 平衡态时, 速度分布在速度空间是各向同性的 (isotropic).

可以证明, 这种情况下, U, V, W 的分布必为正态分布.

证明 由假定, 存在一维密度函数 $p(x)$, U, V, W 均以它为密度函数, 若以 $p(u,v,w)$ 记 (U,V,W) 的联合密度函数, 由假定 (3) 知 $p(u,v,w) = q(u^2+v^2+w^2)$, 由假定 (2)

$$p(u)p(v)p(w) = q(c^2), \quad u^2+v^2+w^2 = c^2 \tag{1}$$

两边取自然对数得

$$\ln p(u) + \ln p(v) + \ln p(w) = \ln q(c^2) \tag{2}$$

两边对 u 求导得

$$\frac{1}{p(u)} \cdot \frac{\mathrm{d}p(u)}{\mathrm{d}u} = \frac{1}{q(c^2)} \cdot \frac{\mathrm{d}q(c^2)}{\mathrm{d}c^2} \cdot \frac{\mathrm{d}c^2}{\mathrm{d}u} \tag{3}$$

整理成

$$\frac{1}{up(u)} \cdot \frac{\mathrm{d}p(u)}{\mathrm{d}u} = \frac{2}{q(c^2)} \cdot \frac{\mathrm{d}q(c^2)}{\mathrm{d}c^2} \tag{4}$$

显然 (4) 的左端与 v, w 无关.

同理可得

$$\frac{1}{vp(v)} \cdot \frac{\mathrm{d}p(v)}{\mathrm{d}v} = \frac{2}{q(c^2)} \cdot \frac{\mathrm{d}q(c^2)}{\mathrm{d}c^2} \tag{5}$$

$$\frac{1}{wp(w)} \cdot \frac{\mathrm{d}p(w)}{\mathrm{d}w} = \frac{2}{q(c^2)} \cdot \frac{\mathrm{d}q(c^2)}{\mathrm{d}c^2} \tag{6}$$

显然 (5) 的左端与 u, w 无关, (6) 的左端与 u, v 无关, 故必有

$$\frac{1}{u} \cdot \frac{\mathrm{d}p(u)}{p(u)\mathrm{d}u} = a \tag{7}$$

其中 a 为常数, 因此

$$p(u)=b\,\mathrm{e}^{\frac{a}{2}u^2} \tag{8}$$

为使 $p(u)$ 为密度函数, 应有 $a<0$, 若记 $a=-\dfrac{m}{kT}$, 其中 m 为质量, k 为玻尔兹曼常数, T 为绝对温度, 则由规范化条件知

$$b=\sqrt{\frac{m}{2\pi kT}} \tag{9}$$

故

$$p(u)=\sqrt{\frac{m}{2\pi kT}}\,\mathrm{e}^{-\frac{mu^2}{2kT}} \tag{10}$$

这正是 $N\left(0,\dfrac{kT}{m}\right)$.

《概率论基础》习题三 36 题以此为出发点导出麦克斯韦分布律.

麦克斯韦的气体动力学理论开创了统计物理学, 连同他在电磁场理论方面的伟大贡献, 使他成为 19 世纪首屈一指的物理学家.

顺便提及, 本札记第一种推导用柯西方程, 第二种推导用求导, 这是解决这类问题的两种常用手段, 一般都可互换.

顺带指出, 第四章 §3 用最大熵概念给出了正态分布的另一个刻画. 此外, 习题三 **43 题也是一种刻画.

教学札记之十五

统计物理学和概率论

一、麦克斯韦分布

由于牛顿力学所取得的成功, 直至 19 世纪上半叶, 物理学的目标是建立以能量守恒定律为统一原理, 充分利用数学工具, 寻求物理现象的力学解释. 例如对热现象就以分子运动学的观点建立了热力学. 它解释了许多现象, 也遗留下一大堆问题.

力学上研究粒子和粒子组的运动, 可以根据初始条件求解其运动方程来进行, 但当一个宏观物体由大量微观粒子组成, 理论上求解数量如此巨大的运动方程和实验上确定这么多粒子的初始条件都是不可能的. 热现象研究的是大数 (10^{23} 数量级) 的微观粒子组成的系统表现出来的宏观现象. 由于粒子之间以及系统与外界之间存在相互作用, 粒子的运动变得十分紊乱, 异常复杂, 因而具有不完全确定性. 一些在单个粒子运动中起重要作用的因素, 像初始条件, 在整个系统状态的演变过程中变得次要甚至不起什么作用了, 而相互作用则导致系统在某一时刻处于何种状态呈现随机性. 系统中除单个粒子遵从力学规律外, 还出现一种与力学规律迥异的新的规律即统计规律性.

1859 年英国物理学家麦克斯韦 (J. C. Maxwell, 1831—1879) 在克劳修斯气体分子运动研究的基础上尝试采用关于速度分布的统计分析方法来研究气体. 麦克斯韦认识到气体的运动太复杂, 以至于难以用牛顿定律去描述. 更重要的是气体的性质例如压强和温度是整体的性质. 一个分子的运动不如作为整体的气体的运动重要, 因此要了解气体的性质就需要把气体作为整体进行概率描述, 而不是用决定论的方式去描述每个分子. 麦克斯韦还证明了在相当简单的假定下, 能导出分子速度的分布, 而系统的宏观实测值则是一切可能微观状态的统计平均值. 于是统计物理学被建立, 它的基本任务就是采用概率论方法探求大数粒子系统的统计规律, 从而找出系统的宏观性质及其规律. 这个理论的诞生对概率论的发展无疑有重大刺激作用.

《概率论基础》用第二章 §4 例 7 (分子运动) 来介绍这类系统, 并在第五章 §1 最后答案中指明宏观结果的稳定性. 习题三 36 题给出气体分子速度的麦克斯韦分布的推导, 而习题四 10 题则求出分子的平均速度和平均动能. 这类结果后为实验所证实.

二、近独立粒子的三种统计

统计物理学研究对象都是大数微观粒子组成的系统. 在承认大数微观粒子系统运动带有统计规律性的前提下, 寻求系统出现在各个微观状态的概率分布, 再利用这个概率分布来求得宏观物理量的统计平均值, 因此确定各微观状态出现的概率是统计物理学的根本问题.

奥地利物理学家玻尔兹曼 (L. Boltzmann, 1844—1906) 发展了最概然 (most probable) 统计法, 它适用于由大数全同和近独立粒子组成的系统. 所谓全同粒子是指质量, 电荷, 自旋, 同位旋等各种内禀固有属性都完全相同的微观粒子. 在经典力学中, 由于可以从粒子运动的不同轨道来分辨不同的粒子, 所以即使是全同粒子也总是可以分辨的, 原则上可将同一种物质的粒子进行编号. 所谓近独立粒子是指粒子之间的相互作用可以忽略, 即粒子之间的相互作用的能量远小于粒子自身的能量, 这种粒子包括理想气体, 黑体辐射中的光子气体等等.

为了描述这种微观粒子, 假定它们在某一时刻总处于相空间的某个点, 称为相点. 随着时间变化它描绘出一条曲线, 称之为相轨迹. 为便于进行统计, 将相空间分成许许多多小体积元, 这种小体积元称为相格. 若将粒子比作球, 则讨论粒子在相格中的分配问题便归结为组合分析中投球入格的计算.

玻尔兹曼早在 19 世纪 70 年代就提出著名的等概率原理: 孤立系统 (它和外界既不交换物质又无相互作用) 处于统计平衡时, 其所有可能的微观状态出现的概率相等. 这是一个合理的假设, 也为大量实验所证实.

这样, 若把 n 个可分辨粒子分配到 g 个相格中则按重复排列共有

$$g^n \tag{1}$$

种方式. 按等概率原理, 它们是等可能的, 从而可以进行各种古典概率计算. 物理学家把这个模型称为**麦克斯韦 – 玻尔兹曼统计**.

这种统计有两大假设: 1° 粒子是可分辨的; 2° 每个相格可容的粒子数无限制.

《概率论基础》第一章 §3 例 4 中提到这种统计.

当研究尺寸为原子大小的物体性质时, 人们发现经典的麦克斯韦 - 玻尔兹曼统计往往与实验结果存在明显的不合. 这样小的物体属于量子力学研究的范畴. 由于微观粒子具有波粒二象性, 每个粒子总是与一个波相联系, 在大量粒子存在的空间中, 会出现两个波重叠在一起的情形, 在波的重叠处, 无法区分第一个粒子的波和第二个粒子的波, 从而不能分辨第一个粒子与第二个粒子, 所以这种全同粒子是不可分辨的, 不能编号的.

1924 年印度物理学家玻色 (Bose) 提出一种新统计: n 个不可分辨的粒子在 g 个无限制相格中的分配共有

$$\binom{n+g-1}{n} \tag{2}$$

种方式. 按等概率原理, 它们是等可能的. 以这个古典概率模型进行概率计算, 物理学家称为**玻色 – 爱因斯坦统计**. 它适用于光子, 自旋量子数为整数的基本粒子及含有偶数个基本粒子的原子或分子或原子核, 通称玻色子.

1925 年泡利 (W. Pauli) 发现 “不相容原理”: 不能有两个或两个以上的全同粒子处在同一状态. 由于这个原理, 在相空间的单位相格 (量子态) 中不能有一个以上的电子, 故电子及其他自旋量子数为半整数的基本粒子在相空间的分布须另考虑.

这时费米 (E. Fermi) 与狄拉克 (P. A. M. Dirac) 创立新的方法: n 个不可分辨的粒子在 g 个满足泡利不相容原理的相格中的分配共有

$$\binom{g}{n} \qquad g \geqslant n \tag{3}$$

种方式. 按等概率原理, 它们是等可能的. 以这个古典概型进行概率计算, 物理学家称为**费米 – 狄拉克统计**. 符合这种统计的粒子称

为费米子.

下面举一个含有 2 个粒子, 3 个量子态的系统作为例子. 可以看出玻尔兹曼系统, 玻色系统和费米系统中 2 个粒子占据 3 个量子态的方式分别为 9, 6, 3.

这段讨论说明, 等概率场合的确定不是纯粹逻辑问题, 经验事实也要考虑. 可能性的假定不是直接可以论证的, 通常必须联系在问题中用此假定的后果的实验验证.

玻尔兹曼系统 以①②各表示一个粒子			玻色系统 让②=①, ①表示一个粒子			费米系统 让②=①, ①表示一个粒子		
量子态			量子态			量子态		
Ⅰ	Ⅱ	Ⅲ	Ⅰ	Ⅱ	Ⅲ	Ⅰ	Ⅱ	Ⅲ
①②			①①			①	①	
	①②			①①			①	①
		①②			①①	①		①
①	②		①	①				
②	①			①	①			
	①	②	①		①			
	②	①						
①		②						
②		①						

三、近独立粒子系统的最概然分布

考虑一个由大数全同近独立粒子组成的孤立系统, 其能量为 E, 体积 V 和粒子数 N 是确定的. 以 $\varepsilon_i(i=1,2,\ldots)$ 表示粒子的能级, g_i 表示能级 ε_i 的简并度 (一个能级的量子态数叫做该能级的简并度). 系统的一个微观态就是 N 个编了号的粒子在能级 $\varepsilon_1,\varepsilon_2,\ldots$ 中各个相格 (量子态) 的一种分配. 对 N 个粒子在各能级上的一个分配, 物理学家以 $\{n_i\}=\{n_1,n_2,\ldots,n_i,\cdots\}$ 表示在能级 ε_i 有 n_i 个粒子, 并称为一个分布. 显然

$$\sum_i n_i = N, \quad \sum_i n_i\varepsilon_i = E \tag{4}$$

一个分布与系统的一个宏观态相对应, 给定一个分布 $\{n_i\}$, 就

确定了占据每个能级的粒子数, 但还不能够确定系统的微观态, 这是因为能级的简并, 每个能级不只有一个量子态, 因此一个分布往往对应许多不同的微观态. 在分布 $\{n_i\}$ 给定后, 如何确定系统的微观态呢? 对玻尔兹曼系统, 还须确定处在各能级 ε_i 上的是哪 n_i 个粒子, 以及每一能级 ε_i 上的 n_i 个粒子占据其 g_i 个量子态的方式.

根据等概率原理, 一个宏观态对应的微观态越多, 这一宏观态出现的概率越大. 在大数粒子系统中系统常处于概率最大的宏观态, 它可近似地看作平稳态, 所以求平稳态的分布就是求微观态最多的分布即最概然分布. 这可用在约束 (4) 下的拉格朗日乘数法来求得.

玻尔兹曼系统

n_i 个粒子占据 g_i 个量子态的方式数为 $g_i^{n_i}$, 任意分布 $\{n_i\}$ 所包含的微观态数为

$$\Omega^{M.B}\big(\{n_i\}\big) = \frac{N!}{\prod\limits_i n_i!}\prod_i g_i^{n_i} \tag{5}$$

用拉格朗日乘数法, 求在 (4) 的约束下的最概然分布, 可从下式出发:

$$\ln\Sigma = \ln\Omega^{M.B}\big(\{n_i\}\big) + \alpha\Big(N - \sum_i n_i\Big) + \beta\Big(E - \sum_i n_i\alpha_i\Big)$$

其中 α, β 为乘数.

利用斯特林公式

$$\ln n! = n\big(\ln n - 1\big)$$

最终求得玻尔兹曼分布

$$n_i^{M.B} = g_i\,\mathrm{e}^{-\alpha-\beta\varepsilon_i} \tag{6}$$

其中 $\beta^{-1} = kT$, 这里的 k 为玻尔兹曼常数, T 为绝对温度.

这个分布适用于可区分的定域粒子, 如理想气体等.

特别地, 理想气体在平衡状态下, 能量 $\varepsilon = \dfrac{1}{2}m\big(v_x^2 + v_y^2 + v_z^2\big)$,

玻尔兹曼分布化为

$$f\left(v_x, v_y, v_z\right) = A\,\mathrm{e}^{-\frac{m(v_x^2+v_y^2+v_z^2)}{2kT}}$$

由规范化知

$$A = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}}$$

$$f\left(v_x, v_y, v_z\right) = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \mathrm{e}^{-\frac{m(v_x^2+v_y^2+v_z^2)}{2kT}}$$

因此, 气体分子的速度向量 $V = (V_x, V_y, V_z)$, 各分量相互独立, 且均服从 $N\left(0, \frac{kT}{m}\right)$, 依《概率论基础》习题三 36 题可知速度 $S = \sqrt{V_x^2 + V_y^2 + V_z^2}$ 服从麦克斯韦分布.

麦克斯韦于 1859 年在 V_x, V_y, V_z 具有连续密度函数, 相互独立, (V_x, V_y, V_z) 的密度函数在 (v_x, v_y, v_z) 点的值仅与 $s = \sqrt{v_x^2 + v_y^2 + v_z^2}$ 有关 (即各向同性) 的假定下导出 V_x, V_y, V_z 均服从正态分布 (参看教学札记之十四 "正态分布的两种刻画"), 并在此基础上导出气体分子的麦克斯韦分布. 这些结果于 1892 年由迈克耳孙 (Michelson) 用气体放射光谱法 (间接) 实验证明, 于 1920 年由施特恩 (O. Stern) 用原子射束方法直接实验证明 (吴大猷《理论物理 (第五册) 热力学, 气体运动论及统计力学》, 北京, 科学出版社, 1983).

玻色系统

每个量子态对占据的粒子数没有限制, n_i 个粒子占据 g_i 个量子态的方式数为 $\binom{g_i + n_i - 1}{n_i}$, 故任一分布 $\{n_i\}$ 所包含的微观态数为

$$\Omega^{B.E}(\{n_i\}) = \prod_i \frac{(g_i + n_i - 1)!}{n_i!(g_i - 1)!} \tag{7}$$

在约束 (4) 下导出最概然分布

$$n_i^{B.E} = \frac{g_i}{\mathrm{e}^{\alpha+\beta\varepsilon_i} - 1} \tag{8}$$

这分布适用于玻色子.

费米系统

每个量子态至多只能有一个粒子占据, 必须 $g_i \geqslant n_i$, n_i 个粒子占据 g_i 个量子态的方式数为 $\binom{g_i}{n_i}$, 故任一分布 $\{n_i\}$ 所包含的微观态数为

$$\Omega^{F.D}(\{n_i\}) = \prod_i \frac{g_i!}{n_i!(g_i - n_i)!} \tag{9}$$

在约束 (4) 下导出最概然分布

$$n_i^{F.D} = \frac{g_i}{e^{\alpha+\beta\varepsilon_i} + 1} \tag{10}$$

这分布适用于费米子.

布里渊方法

布里渊 (L. Brillouin) 采用下面方法把上述三种统计统一起来.

设 n_i 个粒子以下列方式分配于 g_i 个相格: 首个粒子有 g_i 个分配法, 次一个粒子有 $g_i - a$ 个分配法, 第三个粒子有 $g_i - 2a$ 个分配法, $\cdots\cdots$, 第 n_i 个粒子有 $g_i - (n_i - 1)a$ 个分配法. 因此这 n_i 个粒子在 g_i 个相格的分配法为

$$g_i(g_i - a)(g_i - 2a)\cdots\big(g_i - (n_i - 1)a\big)$$

若这 n_i 个粒子是不可分辨的, 则分配法为

$$\frac{g_i(g_i - a)(g_i - 2a)\cdots\big(g_i - (n_i - 1)a\big)}{n_i!}$$

在约束 (4) 下导出最概然分布

$$n_i = \frac{g_i}{e^{\alpha+\beta\varepsilon_i} + a} = \begin{cases} \dfrac{g_i}{e^{\alpha+\beta\varepsilon_i} - 1}, & a = -1, \quad \text{B-E 分布} \\ g_i\, e^{-\alpha-\beta\varepsilon_i}, & a = 0, \quad \text{M-B 分布} \\ \dfrac{g_i}{e^{\alpha+\beta\varepsilon_i} + 1}, & a = 1, \quad \text{F-D 分布} \end{cases}$$

当 $a = 1$ 时, 分配法有不与另一粒子同格之趋势, 此效应与泡

利不相容原理相同; 当 $a=0$ 时, 格中有无粒子不起影响; 当 $a=-1$ 时, 粒子有较大概率聚于一格.

以上分布在现代物理学中的基础地位与众多应用不在此处讨论.

教学札记之十六

量子力学的建立和概率论

W. 汤姆孙 (开尔文勋爵) 在 1901 年指出: 在物理学晴朗的天空中漂浮着两团乌云.

到 19 世纪末, 经典物理学, 主要是经典力学、热力学和经典统计物理、经典电动力学, 已发展得相当完善了, 许多物理现象都能得到定性和定量的解释, 但是也出现一些新实验, 其结果无法在经典物理学中解释.

物理学是实验科学, 新的实验事实提示了原有理论的不足, 也促进了新的理论的诞生. 20 世纪初最为物理学家困惑的实验有两个: 一个是迈克耳孙的观察证实了地球上的光速不依赖于地球绕太阳的运动, 这与传统的 "以太" 学说严重矛盾; 另一个是关于黑体辐射的实验与理论的背离. 这就是开尔文勋爵所说的两团乌云. 但正是对这两个实验事实的研究产生了 20 世纪物理学的两大新理论. 前者是爱因斯坦的相对论, 后者是量子力学, 它们的出现改变了物理学也改变了整个地球的面貌. 下面我们来回顾量子力学的建立过程.

一、黑体辐射与普朗克的量子论

上述黑体辐射的实验得出了辐射能量密度 E_ν 随频率 ν 的分布规律. 1893 年维恩 (Venn) 利用经典热力学和电动力学给出了经验公式

$$E_\nu \mathrm{d}\nu = C_1 \nu^3 \mathrm{e}^{-C_2\nu/T} \mathrm{d}\nu \tag{1}$$

这公式称为**维恩公式**, 只在高频部分与实验数据吻合.

1899 年瑞利 (Rayleigh) 和金斯 (Jeans) 利用经典统计物理学和电动力学推导出

$$E_\nu \mathrm{d}\nu = \frac{8\pi kT}{C^3}\,\nu^2 \mathrm{d}\nu \tag{2}$$

这公式称为**瑞利– 金斯公式**, 其中 C 为光速, 它只在低频区与实验相符, 在高频是发散的.

1900 年普朗克 (M. Planck, 1858—1947) 提出了一个两参数公式,

$$E_\nu \mathrm{d}\nu = \frac{C_1 \nu^3 \mathrm{d}\nu}{\mathrm{e}^{C_2\nu/T}-1} \tag{3}$$

这个公式在高频区化为维恩公式, 而在低频区 ($\mathrm{e}^{C_2\nu/T}-1 \approx C_2\nu/T$) 以瑞利–金斯公式为极限 ($\frac{C_1}{C_2} = \frac{8\pi k}{C^3}$, k 为玻尔兹曼常数). 这个公式在全频段与观察结果惊人地符合.

两个月后, 普朗克发现, 如作假定: 对于一定频率 ν 的辐射, 物体只能以 $h\nu$ 为单位吸收或发射它, 则能从理论上导出**普朗克公式** (3).

这就是普朗克的 "量子说". 它断言, 物体吸收或发射电磁波只能以 "量子" 的方式进行, 每个 "量子" 的能量为

$$\varepsilon = h\nu \tag{4}$$

这里的 h 为一个普适常数, 现称为**普朗克常数**, $h = 6.626 \times 10^{-34}\ \mathrm{J\cdot s}$.

二、光电效应与爱因斯坦的光量子

普朗克的量子说与经典力学中能量连续的概念完全不相容, 因此他的理论并未受到人们的重视. 1905 年年轻的爱因斯坦 (A. Einstein, 1879—1955) 发表了三篇论文, 一篇创立相对论, 一篇解决了布朗运动的难题, 第三篇解释了光电效应. 最后那篇文章使他后来获得诺贝尔奖.

1900 年勒纳 (Lenard, 1862—1947) 的实验证明: 金属在紫外光的照射下发射电子, 这种逸出的电子称为光电子, 光电效应的实验结果是:

(i) 存在临界频率 ν_0, 当入射光的频率 $\nu < \nu_0$ 时, 无光电子逸出. 只有当 $\nu \geqslant \nu_0$ 时, 短时间内就能观察到光电子. 不同的金属有不同的 ν_0.

(ii) 出射的光电子的能量只与入射光的频率 ν 有关, 而与入射光的强度无关.

(iii) 入射光的强度只影响光电流的强弱.

这些实验结果, 特别是 (i) 和 (ii), 无法用经典电磁理论解释, 因为按经典电动力学, 光是电磁波, 电磁波的能量决定于它的强度, 即只与振幅有关而与频率无关, 而要释放光电子则需要足够的能量.

爱因斯坦提出了 "光量子" 概念, 他认为辐射场就是由光量子组成, 每一个光量子的能量 E 与辐射的频率 ν 的关系是

$$E = h\nu \tag{5}$$

他还根据他同年提出的相对论中给出的光的动量和能量的关系 $p=\dfrac{E}{C}$ 提出光量子的动量 p 与辐射的波长 $\lambda\,(=C/\nu)$ 有下列关系:

$$p = \frac{h}{\lambda} \tag{6}$$

(5) 与 (6) 称为**普朗克–爱因斯坦关系**.

采用光量子概念后光电效应中出现的疑难都迎刃而解, 当光照射到金属表面时, 一个光量子的能量可以立刻被金属中的自由电子吸收, 当频率 ν 超过临界频率 ν_0 时光量子的能量足够大, 电子才可能克服脱出功 A 而逸出金属表面, 逸出电子的动能为

$$\frac{1}{2}mv^2 = h\nu - A \tag{7}$$

只与 ν 有关, 而不依赖于照射光的强度.

至此爱因斯坦成功解释了光电效应. 后来他还把能量不连续

的概念应用于固体中原子的振动, 成功解决了比热容在绝对温度 0 K 时趋于 0 的难题.

从此, 普朗克的量子论受到物理学家的注意.

三、原子结构与玻尔的量子论

J. J. 汤姆孙 (Thomson, 1856—1940) 1896 年发现电子后于 1904 年提出过原子模型: 正电荷均匀分布于原子中而电子镶嵌其中. 1911 年卢瑟福 (Rutherford, 1871—1937) 根据 α 粒子对原子散射中出现的大角度偏转现象提出了原子的 "有核模型": 原子的正电荷以及几乎全部质量集中在中心很小区域中形成原子核, 而电子则围绕原子核旋转. 这个模型能解释 α 粒子的大角度偏转, 但遇到两大难题: (i) 原子的大小: 找不到一个合理的特征长度; (ii) 原子的稳定性问题: 电子围绕原子核旋转的运动是加速运动, 按经典电动力学, 电子将不断辐射能量而减速, 轨道半径会不断减小, 最后将掉到原子核上去, 原子随之坍塌. 但事实并非如此.

丹麦年轻物理学家玻尔 (N. Bohr, 1885—1962) 1912 年来到卢瑟福实验室, 对原子模型很感兴趣. 次年又有机会了解到原子线状光谱的规律, 并找到了原子光谱与原子结构之间的本质联系, 终于提出了他的原子的量子论.

玻尔认为量子理论或许有可能作为描写原子结构的基础, 因此也有可能充当原子物理学, 光谱学和化学在逻辑上贯彻一致的理论图景的统一基础. 这个统一图景的中心概念是分立的定态概念.

他假定:

(1) 原子能够而且只能够稳定地存在于与分立的能量 $(E_1, E_2, \cdots)$ 相应的一系列状态中, 这些状态称为定态. 因此原子能量的任何变化, 包括吸收与发射电磁辐射, 都只能在两个定态之间以跃迁的方式进行.

(2) 原子在两个定态 (分别属于能级 E_n 和 E_m, 设 $E_n > E_m$)

跃迁时, 发射或吸收的电磁辐射的频率由下式给出:

$$h\nu = E_n - E_m \text{ (频率条件)} \tag{8}$$

这里提出的定态、量子跃迁、频率条件都是玻尔很了不起的创见.

玻尔再利用对应原理 (大量子数极限下, 量子体系的行为与经典体系相同), 求出了氢原子的能级公式, 得出了可见光谱中的巴耳末 (Balmer) 系, 红外波段的帕邢 (Paschen) 系并预言了紫外波段的莱曼 (Lyman) 系, 获得巨大的成功.

玻尔的理论也有困难, 例如不能解释氦原子等的光谱, 而且只能给出光谱线的频率而不能给出强度等等.

四、波粒二象性与德布罗意波

在光的波粒二象性和玻尔原子论的启发下, 法国物理学家德布罗意 (L. V. de Broglie, 1892—1987) 仔细分析了光的微粒说与波动说的发展历史, 并注意到几何光学与经典粒子力学的相似性, 根据类比的方法, 于 1923 年提出假说: 一切微观粒子都具有波粒二象性. 他的思维逻辑是: 对于电磁波, 人们过去否认其粒子性, 因而在微观领域出现了困难. 当承认其波粒二象性之后, 困难就解决了. 那么, 对实物粒子 ($m \neq 0$), 人们是否存在着相反的认识错误呢? 人们过去从不认为它们还具有波动性, 经典物理学在微观领域遇到的困难的症结是否就在于没有认识到粒子也具有波动性呢?

德布罗意假定, 与一定能量 E 和动量 p 的物质粒子相联系的波 (他称为 "物质波") 的频率和波长分别为

$$\nu = \frac{E}{h} \tag{9}$$

和

$$\lambda = \frac{h}{p} \tag{10}$$

这两个公式称为**德布罗意关系**, 相当于假定普朗克–爱因斯坦关于光子的能量公式 (5) 和动量公式 (6) 也适用于自由粒子. 他提出这

个假定的动机, 一方面是企图把作为物质存在的两种形式 ($m \neq 0$ 的实物粒子和 $m = 0$ 的光) 的理论统一起来, 另一方面是为了更深刻地去理解微观粒子能量的不连续性, 以克服玻尔理论带有人为性质的缺陷.

1927 年戴维逊和革末用一束具有一定能量和动量的电子射向金属镍单晶表面观察到电子衍射的现象, 从而证实了德布罗意关系 $\lambda = h/p$ 是正确的. 后来, 无数的事实都表明, 不仅电子, 而且质子、中子、原子等都具有波动性, 波动性是物质粒子普遍具有的.

仿照经典平面波, 描述自由粒子波粒二象性的平面波, 依 (9), (10) 在 $\boldsymbol{r}$ 处可写成

$$\begin{aligned}\Psi &= A\,\mathrm{e}^{i\,2\pi\left(\frac{\boldsymbol{r}\cdot\boldsymbol{n}}{\lambda}-\nu t\right)} \\ &= A\,\mathrm{e}^{i\frac{2\pi}{h}(\boldsymbol{p}\cdot\boldsymbol{r}-Et)}\end{aligned} \tag{11}$$

其中 $\boldsymbol{n}$ 表示动量 $\boldsymbol{p}$ 的方向.

Ψ 称为**自由粒子波函数**, 只能取复数形式.

五、量子力学的建立

量子力学是在 1923 年到 1927 年这段时间中建立起来的, 两个彼此等价的理论——矩阵力学与波动力学, 几乎同时被提出.

海森伯 (W. Heisenberg, 1901—1976) 于 1925 年提出矩阵力学, 赋予每一个物理量以一个矩阵, 关系式中出现普朗克的作用量子 h, 从宏观角度看 h 是微不足道的, 当 $h \to 0$, 矩阵力学中的量之间的关系将回到经典力学. 矩阵力学成功解释了许多经典物理学的难题获得很大成功.

薛定谔 (E. Schrödinger, 1887—1961) 于 1926 年提出了描述物质波的波动方程, 在波动力学中出现的是二阶偏微分方程, 分立能级的问题表现为一定边界条件下解微分方程的特征值问题. 随后他还证明了波动力学与矩阵力学的等价性.

玻恩 (M. Born, 1882—1970) 于 1926 年给出了波函数的概率

诠释, 提出了概率波的概念.

1927 年海森伯提出测不准关系式.

狄拉克 (P. A. M. Dirac) 于 1927 年提出电磁场的量子理论.

到此, 涉及非相对论性的实物粒子与电磁场作用的问题, 原则上都可以解决.

六、薛定谔的波动方程

对自由粒子的平面波 (11) 式分别对 t 和 x, y, z 求偏导后得

$$i\hbar\frac{\partial\Psi}{\partial t} = E\Psi \tag{12}$$

$$-\hbar^2\nabla^2\Psi = p^2\Psi \tag{13}$$

式中拉普拉斯算子 $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, 而 $\hbar = \frac{h}{2\pi}$.

对自由粒子, $E = \frac{p^2}{2m}$, 故得

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi \tag{14}$$

当粒子处于外力场中, 粒子的位势为 $U(\boldsymbol{r})$ 时, 其波函数 $\Psi(\boldsymbol{r}, t)$ 满足

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + U(\boldsymbol{r})\Psi \tag{15}$$

公式 (15) 称为**薛定谔方程**, 它是量子力学的基本假定之一.

波函数 $\Psi(\boldsymbol{r}, t)$ 是薛定谔方程的解, 可以完全描述粒子的状态. 波函数是复数, 而 $|\Psi|^2$ 具有实际的物理意义. 因此波函数应满足一些带有普遍性的条件, 称为**波函数的标准条件**: 在 $\boldsymbol{r}$ 的变化范围内, $\Psi(\boldsymbol{r}, t)$ 必须有限、单值和对 $\boldsymbol{r}$ 分量的一阶偏导数连续.

可证这种波函数满足波的叠加原理, 这是波的重要特征.

七、波函数的概率诠释

电子究竟是什么东西呢? 是粒子, 还是波? 电子既不是经典粒子也不是经典的波. 也可以说, 电子既是粒子也是波, 它是粒子和

波动两重性的统一，但这个波不是经典意义下的波，粒子也不是经典意义下的粒子.

电子所呈现出来的粒子性只是经典粒子概念中的“原子性”或“颗粒性”，即总是以具有一定的质量和电荷等属性的客体出现在实验中，但不与“粒子有确切的轨道”的概念有什么联系. 而电子呈现出波动性，也只不过是波动最重要的东西——波的叠加性，但并不一定与某种实在的物理量在空间的波动联系在一起.

把粒子性与波动性统一起来的是玻恩. 他在 1926 年提出了**概率波**，把微观粒子的“原子性”和波的“叠加性”统一起来. 他认为德布罗意提出的“物质波”或薛定谔方程中的波函数所描述的，并不像经典波那样代表什么实在物理量的波动，只不过是刻画粒子在空间的概率分布的概率波而已. 微观粒子呈现出来的波动性只是反映微观客体运动的一种统计规律性.

根据波函数的概率诠释，波函数 $\Psi(\boldsymbol{r})$ 满足

$$\int |\Psi(\boldsymbol{r})|^2 \, \mathrm{d}\boldsymbol{r} = 1 \tag{16}$$

的规范性条件.

这样一来，波函数的物理意义是：$|\Psi|^2$ 表示微观粒子在 t 时刻出现在 $\boldsymbol{r}$ 处单位体积内的概率密度. 量子力学在实际中的应用表明，这一解释是正确的.

因此量子力学的波函数满足了同时反映波粒二象性的要求. 一方面，承认粒子是一颗一颗的，具有粒子性，这是粒子的最根本的内容；另一方面，粒子在空间某区域出现的概率随 $|\Psi|^2$ 而波动，可以解释干涉和衍射等波动现象.

八、测不准关系

在经典力学中，用坐标 $\boldsymbol{r}$ 和动量 $\boldsymbol{p}$ 描述粒子的运动状态，其他力学量都是 $\boldsymbol{r}$ 和 $\boldsymbol{p}$ 的函数. 量子力学认为，微观粒子的运动状态必须用波函数 $\Psi(\boldsymbol{r}, t)$ 描述，$\boldsymbol{r}$ 和 $\boldsymbol{p}$ 只能描述经典粒子的粒子性，

不能描述波粒二象性.

如果一定要同时用 $\boldsymbol{r}$ 和 $\boldsymbol{p}$ 描述微观粒子, 则所得到的 $\boldsymbol{r}$ 和 $\boldsymbol{p}$ 不能同时有确定值. 这是 1927 年海森伯提出的**测不准关系式**

$$\Delta x \Delta p_x \geqslant \frac{\hbar}{2} \tag{17}$$

其中 Δx 表示坐标的不确定程度, Δp_x 表示动量在 x 方向的分量的不确定程度.

测不准关系可理论证明, 也有大量实验证明. 测不准关系式指出了经典理论的适用范围. 测不准关系式也否定了微观粒子的 "轨道运动". 它也指明势能和动能不可能同时有确定值.

量子力学表明微观物理世界的不确定性和使用概率概念的必要性, 它扩大了概率论的应用领域, 提升了概率论在科学中的地位, 也极大地推动了概率学科的发展.

教学札记之十七

随　机　数

一、何谓随机数

袋中放有 10 只外形质地等等完全相同的球, 分别写上 0, 1, 2, $\cdots$, 9 等十个数字. 从中摸出一球, 记下它上面的数字之后放回, 经充分混合后再摸一球记下数字之后又放回, $\cdots\cdots$, 依次进行下去, 将记录下一串数字, 这就是随机数.

随机数可以看作服从下列分布列的随机变量 ξ 的独立同分布观察值:

ξ	0	1	2	$\cdots$	9
P	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\cdots$	$\frac{1}{10}$

这个分布列称为从 0 到 9 的离散均匀分布.

通常把若干个随机数排成一组, 若干组排成一行, 若干行排成一页, 若干页构成一张表, 这就是随机数表.

若把 5 个随机数排成一组并看作一个数, 则它可以看作另一个随机变量 η 的独立同分布观察值, 这里的 η 服从 00000 到 99999 的离散均匀分布.

如果我们设想把小数点放在上述随机数的前面, 又取足够多的位数以保证精度, 则得到的随机数便可以看作服从 $U(0,1)$ 的随机变量 ζ 的独立同分布观察值. 口头上或书面上出现的随机数多半就是这个含义, 较准确的应称为 $(0,1)$ 均匀分布随机数.

上述做法太麻烦, 不能满足实际工作对随机数的大量需求, 因此就得另想产生随机数的新方法, 不外数学与物理两种途径, 近代则主要利用计算机.

二、随机数的使用

随机数的系统使用恐怕不到一百年. 最早使用随机数的领域大概是抽样调查与试验设计, 基本上都跟统计关系密切.

1927 年英国统计学家 Tippert 的《随机抽样数》出版, 该书含 41600 个随机数, 排列成 4 个一组, 每页有数行, 计有 26 页, 据说这些数字是作者从某些英国社会调查报告的数字中截头去尾拼接起来的. 随后费希尔 (Fisher) 和耶茨 (Yates) 的统计表问世, 其中有一个随机数表, 含 15000 个随机数, 是由 20 位对数表截取第 15—19 位排列而成的.

英国统计学家卡尔・皮尔逊是最早利用随机数的人之一, 他还指导学生们利用随机数计算在求统计量分布中遇到的重积分, 得出很多结果, 也是他鼓励 Tippert 搞出上述随机数表.

蒙特卡罗方法的出现使随机数的使用进入一个全新的阶段. 蒙特卡罗方法是 20 世纪 40 年代美国科学家在研制核武器的工作中为解决在瞬间内发生的复杂的物理过程的数值计算问题而发展起来的. 战后随着计算机技术的迅速发展, 蒙特卡罗方法也发展成为

一种功能独特的计算技术, 可用来解决各门学科中的形形式式问题. 该应用大体可分为三种形式: 直接模拟, 积分计算, 动态模拟即马尔可夫链蒙特卡罗 (MCMC), 都需要产生大量随机数.

三、随机数的产生

一个随机数是指某些随机变量所取的某一个特定值, 严格讲它只能用某些随机物理过程来产生. 理论上不成问题, 实际做起来困难很多. 因此实际使用的都是 "准随机数" 或 "伪随机数".

伪随机数是利用某些数学公式计算而产生的数列, 它们能够通过随机数的一系列统计检验, 主要是分布检验与独立性检验. 统计学家早就提出基于概率原理的各式各样统计检验法, 只要适当选用就可以.

对于伪随机数产生器的基本要求是: 良好的统计分布特性, 高效率的伪随机数产生, 伪随机数产生的循环周期长, 产生程序可移植性好 (譬如可在不同型号计算机上实现, 与字长无关等). 要求有良好的统计分布特性就是要能通过上述一系列统计检验, 这当然是最重要的. 现有的产生程序还没有一个是十全十美的, 但许多都还能用, 总之这也有一个不断探索、实践、创新的过程.

最早的 $U(0,1)$ 伪随机数产生器可能是 "平方取中器", 使用效果不太理想.

如今比较流行的是同余产生器:

$$x_{n+1} = (ax_n + c)(\mathrm{mod} m)$$

$$\xi_n = \frac{x_n}{m}$$

其中 x_0 称为种子 (或初值), a 称为乘子, c 称为增量, m 称为模, 都是大于 0 的整数.

用这种算法到某一次一定会得到以前出现过的数, 这时序列进入循环, 研究的目标之一是要找出尽可能长的循环周期, 这就要求仔细挑选模 m 和乘子 a 的最佳搭配. 这时数论会有些帮助.

$c=0$ 的情况称为乘同余法, 周期可能短些, 但随机性会较好, 因此可能用得更多. 这时的递推公式为

$$x_{n+1}=ax_n(\mathrm{mod}m)$$

$$\xi_n=\frac{x_n}{m}$$

一个被推荐的乘同余随机数产生器是: 阿伦斯 (J. H. Ahrens) 与迪特 (V. Dieter) 提出的

$$x_{n+1}=\rho x_n(\mathrm{mod}m)$$

$\rho=663\ 608\ 941$, x_0 为任何介于 1 到 2^{32} 之间的奇数, $m=2^{32}$.

摘自 Dudewicz & Ralley. The Handbook of Random Number Generation and Testing with TESTRAND Computer Code. American Sciences Press. Columbus, Ohio, 1981.

现在几乎任何一台计算机及其程序系统中都含有 $U(0,1)$ 随机数产生器.

四、任意分布随机数的产生

应用中当然不止需要 $U(0,1)$ 随机数, 而是根据问题的性质需要产生各种分布的随机数. 幸而这里有一个统一的处理原则, 而 $(0,1)$ 均匀分布是核心.

正如《概率论基础》第三章 §3 关于均匀分布的特殊地位中所指出的, 若 $\theta\sim U(0,1)$, 则

$$\xi=F^{-1}(\theta)\sim F(x)$$

其中 $F(x)$ 是任一分布函数, 而 $F^{-1}(y)$ 是它的反函数.

因此只要能产生 $U(0,1)$ 随机数, 用这种逆变换法就能产生任意分布的随机数.

例如, 对指数分布 Exp (λ),

$$F(x)=1-\mathrm{e}^{-\lambda x},\quad x\geqslant 0$$

这时

$$F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y), \quad 0 \leqslant y \leqslant 1$$

因此

$$\xi = -\frac{1}{\lambda}\ln(1-\theta)$$

由于 $1-\theta$ 与 θ 同样服从 $U(0,1)$, 还可以进一步简化为

$$\xi = -\frac{1}{\lambda}\ln\theta$$

这个 ξ 服从 Exp (λ).

因此如果我们用乘同余法产生 N 个 $U(0,1)$ 随机数 $\theta_1, \theta_2, \cdots, \theta_N$, 则利用上式便可产生 ξ_1, $\xi_2, \cdots, \xi_N$, 它们相互独立, 均服从 Exp (λ) 分布, 从而是 Exp(λ) 的随机数.

类似地, 瑞利分布也可用此法处理,《概率论基础》(3.3.42) 式正是这样做的.

不过由于有些分布函数的反函数没有简洁的解析表达式或计算过于繁复, 需要发明其他更有效的随机数产生器. 正态分布就是一例. 因此《概率论基础》中介绍的由博克斯 - 米勒 (Box-Müller) 提出的变换法 (3.3.42) 及中心极限定理法 (5.3.15), 效率更高.

总的说, 对于常用的分布都已经有了特定的比较有效的随机数产生器, 许多做法还跟抽样方法结合起来. 这又是一个快速更新的领域, 因此有兴趣或有需要的读者应参阅最新的书刊.

教学札记之十八

Γ 分 布

一、前言

Γ 分布 $\Gamma(r,\lambda)$ 具有良好的性质和广泛应用, 在概率论中占有重要的地位.

在皮尔逊 1894 年依微分方程导出的频率曲线族中已有 Γ 分布, 属于Ⅲ型. 这个分布有形状参数 r 和尺度参数 λ 可调节, 能用来拟合各类实际问题提供的实测数据. 例如在水文学中, 皮尔逊Ⅲ型曲线常用来拟合洪峰流量的历史数据, 再作外推以估算 "百年一遇" 的最大流量, 为工程设计提供依据.

埃尔朗在电话系统设计的研究中也曾巧妙地利用了形状参数为正整数的 Γ 分布的一个性质突破难关, 因此这种分布被命名为埃尔朗分布. 这事件的梗概简介如下: 实测表明, 电话呼叫流很好地符合泊松过程, 若再假定每次的通话时间服从指数分布 (话务研究在大量场合也证实这点), 埃尔朗建立的电话交换模型导致马尔可夫过程, 从而能导出排队长度、等待时间等重要变量的分布, 问题得以顺利解决. 但是若去掉通话时间为指数分布的假定, 没有了 "无记忆性", 问题难度立即大增. 这时候埃尔朗想出用带正整数形状参数的 Γ 分布来描述通话时间, 一是它可以表为若干个独立指数分布变量之和, 从而化为多阶段服务而每阶段都保持马尔可夫性; 二是该分布可以拟合各种通话时间数据使模型的适用性增加. 这样一来问题得到完满解决. 在这类开创性的工作中埃尔朗成了排队论的先驱.

当然人们对 Γ 分布有深刻认识是在对泊松过程作深入研究后发现埃尔朗分布可用来描述第 r 个到达时刻, 它是前 r 个来到间隔 (服从指数分布) 的独立和.

二、埃尔朗分布与指数分布

若以 $W_1, W_2, \cdots, W_n$ 表示参数为 λt 的泊松过程的第 $1, 2, \cdots, n$ 个到达时刻, 记

$$\tau_1 = W_1$$

$$\tau_2 = W_2 - W_1$$

$$\cdots\cdots\cdots\cdots$$

$$\tau_n = W_n - W_{n-1}$$

则 $\tau_1, \tau_2, \cdots, \tau_n$ 称为泊松过程的第 $1, 2, \cdots, n$ 个来到间隔.

《概率论基础》第三章 §1 已导出 $W_r, r = 1, 2, \cdots$, 服从参数为 r 的埃尔朗分布, 特别地, $W_1 = \tau_1$ 服从指数分布, 书中还指出

$$W_n = \tau_1 + \tau_2 + \cdots + \tau_n$$

其中 $\tau_1, \tau_2, \cdots, \tau_n$ 相互独立, 均服从指数分布 Exp (λ). 这个结论的证明见教学札记之十二 "关于两个等待时间分布序列".

三、正态分布, χ^2 分布与 Γ 分布

《概率论基础》的正文与习题对上述三种分布的关系有许多讨论, 要点有:

(1) 标准正态变量的平方服从 χ_1^2 分布, 见第三章 §3 例 3.

(2) χ_n^2 分布即为 $\Gamma\left(\dfrac{n}{2}, \dfrac{1}{2}\right)$, 见 (3.3.11).

(3) n 个相互独立标准正态变量的平方和服从 χ_n^2 分布, 见习题三 *37 题.

(4) χ^2 分布具有可加性 (3.3.38) 即再生性 (4.5.29).

(5) 多元正态分布与 χ^2 分布的关系见 (4.6.21).

四、随机变量的函数的独立性之一

关于 Γ 分布, 下面是一个刻画性的性质.

命题一 若 ξ 与 η 相互独立, 且 $\xi \sim \Gamma(r_1, \lambda)$, $\eta \sim \Gamma(r_2, \lambda)$, 则 $\xi + \eta$ 与 $\dfrac{\xi}{\eta}$ 独立.

由 Γ 分布的可加性知 $\xi + \eta \sim \Gamma(r_1 + r_2, \lambda)$, 而 $\dfrac{\xi}{\eta}$ 服从 Fisher 的 Z 分布 $Z(r_1, r_2)$, 这是数理统计中一类重要分布.

由于 Exp $(\lambda) = \Gamma(1, \lambda)$, 因此有如下结论:

结论 1 若 ξ 与 η 相互独立, 均服从 Exp (λ), 则 $\xi + \eta$ 与 $\dfrac{\xi}{\eta}$ 相互独立.

这个结论常以例题或习题的形式出现.

由于 $\chi_n^2=\Gamma\left(\dfrac{n}{2},\dfrac{1}{2}\right)$, 又有如下结论:

结论 2 若 ξ 与 η 相互独立, 分别服从参数为 m 和 n 的 χ^2 分布, 则 $\xi+\eta$ 与 $\dfrac{\xi}{\eta}\cdot\dfrac{n}{m}$ 相互独立, $\xi+\eta\sim\chi_{m+n}^2$, 进一步可导出 $\dfrac{\xi}{\eta}\cdot\dfrac{n}{m}$ 服从 F 分布. 见《概率论基础》第三章§3 例 7.

以 Γ 分布为参照, 以 χ^2 分布为中介, 可发现对正态分布也有相似的结论 (虽然它不能由命题一直接推出):

结论 3 若标准正态变量 ξ 与 η 相互独立, 则 $\xi^2+\eta^2$ 与 $\dfrac{\xi}{\eta}$ 相互独立, 前者服从 χ_2^2 即 $\mathrm{Exp}\left(\dfrac{1}{2}\right)$ 分布, 后者服从柯西分布. 见《概率论基础》习题三 39 题.

下面是结论 3 的一种更常见变体.

结论 4 若标准正态变量 ξ 与 η 相互独立, 则 $\sqrt{\xi^2+\eta^2}$ 与 $\arctan\left(\dfrac{\eta}{\xi}\right)$ 相互独立, 前者服从瑞利分布, 后者服从 $U[0,2\pi]$ 分布. 见《概率论基础》第三章§3 例 9, 这是一个有名又有用的结果.

顺便指出, 结论 3 与结论 4 也建立了 χ_2^2 即 $\mathrm{Exp}\left(\dfrac{1}{2}\right)$ 分布与瑞利分布的关系, 而柯西分布与均匀分布的关系则在《概率论基础》第三章§3 例 5 中早已建立.

五、随机变量的函数的独立性之二

下面是关于 Γ 分布的另一个重要性质.

命题二 若 ξ 与 η 相互独立, 且 $\xi\sim\Gamma(r_1,\lambda),\eta\sim\Gamma(r_2,\lambda)$, 则 $\xi+\eta$ 与 $\dfrac{\xi}{\xi+\eta}$ 独立. 前者服从 $\Gamma(r_1+r_2,\lambda)$, 后者服从 $\boldsymbol{\beta}$ 分布 $B(r_1,r_2)$. 见《概率论基础》习题三 38 题.

对照命题一及命题二不难看出 $\boldsymbol{\beta}$ 分布与 Z 分布有着直接的联

系, 目前流行使用 $\boldsymbol{\beta}$ 分布而少用 Z 分布, 因此后者逐渐从教科书中淡出.

由于 $\text{Exp}(\lambda)=\Gamma(1,\lambda)$, 因此有如下结论:

结论 5 若 ξ 与 η 相互独立, 均服从 $\text{Exp}(\lambda)$, 则 $\xi+\eta$ 与 $\dfrac{\xi}{\xi+\eta}$ 相互独立, 且 $\xi+\eta\sim\Gamma(2,\lambda)$, 而 $\dfrac{\xi}{\xi+\eta}\sim U[0,1]$.

再由于 $\chi_n^2=\Gamma\left(\dfrac{n}{2},\dfrac{1}{2}\right)$, 又有

结论 6 若 ξ 与 η 相互独立, 分别服从参数为 m 和 n 的 χ^2 分布, 则 $\xi+\eta$ 与 $\dfrac{\xi}{\xi+\eta}$ 相互独立.

利用 χ^2 分布作中介, 也可建立正态分布的如下结论:

结论 7 若标准正态变量 ξ 与 η 相互独立, 则 $\xi^2+\eta^2$ 与 $\dfrac{\xi^2}{\xi^2+\eta^2}$ 相互独立, 前者服从 $\chi_2^2=\text{Exp}\left(\dfrac{1}{2}\right)$ 分布, 而 $\dfrac{\xi^2}{\xi^2+\eta^2}$ 服从 $B\left(\dfrac{1}{2},\dfrac{1}{2}\right)$ 分布, 其密度函数为

$$p(y)=\frac{1}{\pi\sqrt{y(1-y)}},\quad 0\leqslant y\leqslant 1$$

这是概率论中有名的**反正弦分布**.

由于 Γ 分布具有良好的性质, 它又能包含很多有名分布作为特例, 再加上有广泛的应用, 因此目前它已成为概率论中仅次于正态分布、泊松分布与二项分布之后的第四个重要分布.

教学札记之十九

概率论公理化结构与测度论

1933 年科尔莫戈罗夫提出他的概率论公理化结构, 虽然之前与之后都有人提出过概率论的其他公理化方案, 但是时到今日, 唯有科尔莫戈罗夫的公理化结构为一切数学家所公认, 并写入各种

正式教科书. 因为它实在太杰出了, 无与伦比, 不但给所有概率论概念予严格的定义, 而且满足了各个学科对它的要求, 在后来几十年发展中对概率论提出的新问题也都能在这个框架中进行研究. 因此目前只要提起公理化概率论, 都是指这个结构. 《概率论基础》当然也不例外.

科尔莫戈罗夫概率论公理化结构中, 把事件定义成 σ 域 $\mathscr{F}$ 中的元素, 而概率 P 则定义为 $\mathscr{F}$ 上的规范化测度, 从此概率论就与测度论紧密相关, 如今世界各大学修读概率论的研究生恐怕很少有不读测度论的.

因此要写一本理论上严谨的基础概率论教科书, 想完全摆脱测度论是十分困难的, 但是如果学概率论一定要先学测度论, 或者把测度论的许多内容放在概率论中来讲, 则会把大部分想学概率论的人排斥在外, 而且用这种方式讲基础概率论也不大会成功. 问题还是度的把握.

《概率论基础》对这个难题的处理算是一种尝试. 基本态度是不回避测度论, 但也不炫耀测度论.

《概率论基础》中概率论的基本概念如事件、概率、随机变量、分布、独立性、数学期望、收敛性等都有严格的定义, 大部分命题都能讲得清楚, 但是还剩下 4 处引用测度论结论而未加证明.

为了做到这点, 难点在事件域与随机变量的可测性. 因为按《概率论基础》的体系, 概率是经频率、古典概型、几何概率导入的, 它的三公理的提出, 水到渠成.

事件域 $\mathscr{F}$ 是科尔莫戈罗夫公理化结构的核心, 事实上把事件看作集合, 把概率看作测度在之前的公理化尝试中早就有人提出, 但缺乏事件域 $\mathscr{F}$ 并不成功. 从当代随机过程理论中对事件信息流的倚重可见这种设计的高明, 也是其他公理化方案所无法比拟的.

在基础概率论中, 事件域的落实主要是 $\mathbf{R}^n$ 中的博雷尔点集, 这是学生们在微积分和数学分析中所未遇到的, 有点难度, 经适当讲解也能接受.

《概率论基础》对这些都不回避, 但只用细线条进行, 对读者不苛求, 也让教师能根据具体情况加以发挥.

《概率论基础》中利用测度论结果而未加证明的 4 个地方指:

1. (3.1.2) $\Longrightarrow$ (3.1.1);

2. (3.2.2) $\Longrightarrow$ (3.2.3);

3. (3.2.37) 的证明;

4. 定理 4.1.1 即佚名统计学家公式 (4.1.17) 的证明.

关于 1,《概率论基础》配置的习题三 *49,** 50 题就是用来证明这个结论的, 用的是测度论的标准方法. 要学生自己想到这个证法是相当不容易的, 但是写成习题的这种形式并加了提示, 一般学生还是做得出的.

关于 2, 只是把 1 的一维问题改成 n 维问题, 几何上有差异, 但从测度论观点看则无本质不同, 因此上述证法只要稍加修改就能证出这一结论.

1 的证明已作为习题解答详细写出, 2 就留给读者作为练习.

关于 3, 对二维场合, 要求证明: 若 ξ 与 η 独立, 则

$$P\{\xi \in A,\, \eta \in B\} = P\{\xi \in A\}P\{\eta \in B\}$$

对一切 $A \in \mathscr{B}_1$, $B \in \mathscr{B}_1$ 成立.

它的证明要困难得多. 证明思路是先固定 $B = (-\infty, y)$, 对 A 用上述测度论标准方法一次, 说明对一切 $A \in \mathscr{B}_1$ 成立; 再固定 $A \in \mathscr{B}_1$, 对 B 也用一次标准方法, 从而证明对 $B \in \mathscr{B}_1$, 该式成立. 这个思路不是一般人所能想到的.

进一步的困难是在第二次, 即对 B 用标准方法验证 σ 域的条件时, 发现对相容事件的并, 验证无法进行, 遇到习题二 18 题所碰到的问题. 幸而对不相容事件的并还容易通过, 因此终于找到更巧妙的方法绕过这个困难. 完整写出证明过程篇幅过长, 因此就此打住.

至于 4, 课文已经指出: 在离散型场合, 佚名统计学家公式

(4.1.17) 化为 (4.1.18)

$$Eg(\xi)=\sum_{i=1}^{\infty}g(x_i)p(x_i)$$

只要写出 $g(\xi)$ 的分布就能得到.

而当 ξ 具有密度函数 $p_\xi(x)$ 时, (4.1.17) 化为 (4.1.20)

$$E\eta=Eg(\xi)=\int_{-\infty}^{\infty}g(x)p_\xi(x)\,\mathrm{d}x$$

在 $y=g(x)$ 适合变换法时 (参看《概率论基础》(3.3.12) 及 (3.3.14)) 也可直接证明.

事实上, 这时若令 $y=g(x)$, $x=g^{-1}(y)$, 则 $\mathrm{d}x=\left|[g^{-1}(y)]'\right|\mathrm{d}y$, 因而有

$$\begin{aligned}Eg(\xi)=E\eta&=\int_{-\infty}^{\infty}yp_\eta(y)\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}yp_\xi[g^{-1}(y)]\left|[g^{-1}(y)]'\right|\mathrm{d}y\\&=\int_{-\infty}^{\infty}g(x)p_\xi(x)\,\mathrm{d}x\end{aligned}$$

可见这时 (4.1.17) 只是积分的变量变换公式.

因此对离散型和连续型及其线性组合, 佚名统计学家公式都是可以在微积分的基础上证明的. 不过由于《概率论基础》的数学期望是用斯蒂尔切斯 (Stieltjes) 积分定义的, 因此对一般分布函数, 还是有必要对斯蒂尔切斯 (Stieltjes) 积分证明积分的变量变换公式 (4.1.17), 或在更宽广的测度论的基础上证明它.

科尔莫戈罗夫公理化体系还有其他成功之处, 例如:

(1) 对条件概率与条件数学期望给出一种数学上严格的一般性定义, 使能处理复杂的场合, 例如 $P(A|B)$ 当 $P(B)=0$ 的情况等.

(2) 证明只要有限维分布满足相容性就能对随机序列建立适当的概率空间, 这为一般随机过程论及函数空间测度研究提供了必要的框架.

顺带说明, 研究实值随机变量 ξ 时, 使用的概率空间实际上是 $(\mathbf{R}^1, \mathscr{B}_1, P_1(\cdot))$, 其中概率测度 $P_1(\cdot)$ 由分布函数 $F_\xi(x)$ 导出. 这里要用到测度扩张定理: 由分布函数 $F_\xi(x) = P\{\xi < x\}$ 可以确定概率分布 $P_1(B) = P\{\xi \in B\}, B \in \mathscr{B}_1$.

类似地, 对 n 维随机向量 $(\xi_1, \xi_2, \cdots, \xi_n)$, 由 n 元分布函数 $F(x_1, x_2, \cdots, x_n)$ 可以确定概率分布 $P_n(B), B \in \mathscr{B}_n$.

测度扩张定理也在测度论中证明. 《概率论基础》中没有明显涉及这些内容.

第四章　数字特征与特征函数

章前引言

本章引入若干数字特征，它们以一个数值而刻画了分布的某一特征，在概率论及其应用中十分重要. 最主要的数字特征是数学期望，方差和相关系数，它们都有深刻的概率涵义，书中用两组应用实例从多个角度来加以阐述. 上述数字特征可用矩来概括.

熵与信息量是信息论的核心概念，这里作了简介.

特征函数 (母函数作为其前导) 与分布函数一一对应而具有更好的数学性质，是数学家们经过长期筛选才找到的处理概率问题的有力工具.

以特征函数为工具对多元正态分布作了全面的处理，初步显示特征函数的威力，所证多元正态分布的诸多性质，可说是概率论的理论和应用中最精彩的内容之一.

课文导读

§ 4.1　数学期望

用平均值来标志涨落水平，以中心化来对抗离散倾向，这是人类在长期实践中产生的智慧. 概率论用数学期望来表征随机变量 (或分布函数) 就是一个很恰当的例证.

学习本节，首先必须注意数学期望是如何定义的. 书中分三步走，分别对离散型，连续型，一般随机变量定义数学期望. 在离散

型场合, 数学期望是随机变量的可能取值以取该值的概率为权作加权平均, 有很明显的概率意义和直观解释. 连续型受离散型逼近启发, 数学期望定义为关于密度函数的一个积分. 若把概率看作质量, 则数学期望表征其重心. 对一般随机变量也用离散型逼近, 定义为对于分布函数的一个斯蒂尔切斯积分. 《概率论基础》 采用承认斯蒂尔切斯积分的若干性质, 并说明这种积分的确能包含离散型和连续型为特例, 用来定义一般随机变量的数学期望. 这样做可把从积分论 (在概率论教育中归入测度论, 是其中的一章) 节省下来的时间用来讲更多的概率论.

其次, 应记住重要分布的数学期望, 并注意: (1) 它常有很直观的解释; (2) 它与分布的参数关系密切; (3) 不少分布由它的数学期望唯一决定.

第三, 结合实际问题, 认识数学期望的直观意义, 并进一步理解引进数学期望这一概念对概率论的重要性. 正文 (尤其是应用实例) 和习题都提供了不少典型问题, 更希望读者举一反三.

第四, 要特别关心和注意 "佚名统计学家公式", 这公式可能是概率论中用得最多的公式, 但大部分都在无意识中使用. 有不少人甚至不知道有这么一个公式, 有些人则认为理所当然成立.

假如你还不体会 "佚名统计学家公式" 的重要性, 那么就请你证明数学期望的如下十分直观的性质:

$$E(\xi+\eta)=E\xi+E\eta$$

没有 "佚名统计学家公式" (这里要用的是二维公式), 还真不知道要如何证明它!

上述性质称为数学期望的可加性 (或线性性), 是数学期望最重要的性质, 数学期望因它而得以在若干表征中心趋势的备择者中脱颖而出. 许多讨论因利用它而便捷, 不少习题的解答因它而简化.

§ 4.2 方差, 相关系数, 矩

本节引进随机变量与分布函数的许多数字特征. 其中方差表

征随机变量取值的离散程度, 它与数学期望是最重要的两个数字特征, 它们可以说是 "天作之合", 联手构成 "均值-方差理论", 通常就能对随机问题有相当深刻的描述; 相关系数给出两个随机变量的线性联系程度; 矩和分位数也表征了分布的重要特征.

数学期望概念在概率论早期就产生, 但方差等概念却出现得相当晚, 跟统计学关系也更密切.

对表征随机变量取值与其中心的偏离程度, 方差是优胜者, 主要是因为它便于数学处理, 背后有很强劲的数学理论支持. 但它的竞争者平均绝对偏差也有优点与潜力. 想深入了解的读者可参看作者的另一本书 (参考书目[2]) 的第七章与第九章.

学习本节, 首先应记住方差的定义和它的计算公式 (4.2.1), 还要记住重要分布的方差, 并注意在它们的计算中一再使用佚名统计学家公式.

其次, 要记住方差的主要性质. 性质 4 (4.2.8) 是联系数学期望与方差的一个不等式, 有重要后续应用.

《概率论基础》为严格证明 "方差为 0 的随机变量是常数" 这一论断, 提早引进了切比雪夫不等式. 切比雪夫不等式的发现是概率论历史上的一个里程碑, 从此俄罗斯学派在概率论中独领风骚, 古典型框框被突破, 随机变量, 数学期望成为主要语言. 关于该不等式在下章有更多叙述.

相关系数 ρ 的引入该归功于统计学. 从此对两个随机变量的联系程度中最简单的一种——线性联系程度有了一个数量指标. 不相关性大大弱于独立性, 但等价于方差的可加性 $D(\xi+\eta)=D\xi+D\eta$ 与数学期望的可乘性 $E\xi\eta=E\xi\cdot E\eta$, 因此很重要. 协方差的引进让方差阵成形, 它与一元时的方差一样, 成为多元场合的基本参量. 《概率论基础》对相关系数的处理特别详尽, 有异于别的教本. 例如证明了相关系数在线性变换下的不变性; 详细讨论了不相关与独立性的关系; 从正态场合的特殊性讲到边际分布为正态而联合分布非正态之例; 由二值变量的特殊性引入事件的相关性 (参看教

学札记之二十三 "事件的独立性与相关性"), 还顺便讨论了抽样调查中的一个核心问题.

希望四个应用实例让读者对均值 - 方差理论留下较深印象, 而条件数学期望, 最佳线性预测一段则展示了这个理论的广阔前景.

矩在极限理论和数理统计中都是最重要的工具之一, 是概括性很强的一类数字特征, 而分位数则提供了分布的另一类数字特征.

顺带指出, 在研究一个变量与另一变量的线性联系程度时, 因为要考虑这两个变量变化是同向变化还是反向变化, 因此具有符号的相关系数是主要指标, 但是当推广到多个变量时, 诸变量的变化方向无法统一, 这时通常用决定系数 (coefficient of determination) ρ^2, 它才真正在线性变换下保持数值不变. 在预测误差和回归中, 真正出现的也是 ρ^2, 见《概率论基础》(4.2.65).

§ 4.3 熵与信息

本节给出信息论的一个导引, 在概率论中讲信息论是少数派, 但也有若干好处.

为导出随机试验的不肯定性的度量, 依照香农的思路先提出三条基本要求, 从而导出表达式 (4.3.5), 因它与热力学中的熵表达式相同, 也称之为熵.

这样导出的熵的一些基本性质被证明. 人们发现, 用它作为不肯定性程度的度量是符合要求的. 在此基础上, 引进了信息量的概念并建立了信息论, 称为香农信息论, 这是迄今为止最通行的信息理论.

连续型分布的熵被平行定义, 也证得它的若干基本性质, 并以之为工具证明了三种重要分布——均匀分布, 指数分布, 正态分布的特殊地位, 即在不同的约束下具有最大熵.

本节与全书完全独立, 是否列入讲授大纲, 可由任课教师自主决定.

§ 4.4 母函数

母函数显示出因有力分析工具的导入使得原本困难的概率问题得到容易的解决.

母函数主要用于古典问题或离散模型. 雅科布·伯努利制成一份等同于如下定理的表格: 投掷 n 颗骰子能得到 m 点的方法数等于表达式

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^n$$

中 x^m 的系数. 这可以看作是母函数法的肇始 (参看《概率论基础》第四章 §4 例 9).

现代意义上, 母函数是概率分布的一种变换. 它把一串概率分布变成一个函数, 而且是性质良好的幂级数, 很便于分析处理. 母函数犹如一只袋子把许多珠宝 (概率分布值) 放在一起, 带进带出, 随时取用, 十分方便.

母函数的优良性质与用途包括:

(1) 与概率分布一一对应, 包含概率分布的全部信息, 不像数字特征只保留部分信息;

(2) 便于求数字特征. 若同时求一个分布的多个数字特征, 则经由母函数简直是最速途径;

(3) 处理独立和的最好工具——把卷积化为乘积;

(4) 特别适合于处理随机个独立随机变量之和, 参看教学札记之二十四 "母函数与分支过程";

(5) 用于证明若干离散型分布的再生性.

在《概率论基础》中, 母函数是作为下节要讲的特征函数的先导而安排的, 教学中也可选择跳过.

§ 4.5 特征函数

概率论在研究极限定理过程中寻找有力的分析工具, 最后找到特征函数有它的必然性. 特征函数是分布函数的傅里叶–斯蒂尔切斯变换, 而傅里叶变换是分析中最有力的工具.

有个有趣的讲法说，有人问英国人，如果要把所有书籍全部烧光只留一部书，要留哪部？英国人答道要留莎士比亚全集．假如向数学家提出同类问题，只许保留一种工具，那么答案大概是傅里叶分析．

傅里叶分析应用于许多数学分支，但着重点并不完全一样．在概率论中最起作用的是一个不很起眼的性质：傅里叶变换把卷积化为乘积．出发点是研究独立随机变量和的极限定理．从分布出发，正如《概率论基础》第三章 §3 指出的，遇到的是分布的卷积，简直无从着手，但一用傅里叶变换化为特征函数，则是再简单不过的乘积．一如从地面上走，要跨越崇山峻岭，大河险滩，步履维艰，为达目的地只能找航空器，解决升空，飞行，导航，降落等问题．本节就是做这类基础工作的．

本质上，傅里叶分析涉及复变函数论，但在特征函数中，若善用欧拉公式 $\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$，则大部分问题都化为实的函数．经仔细处理，《概率论基础》(包括习题解答) 基本上不用复变函数论，就能在微积分基础上顺利进行．

本节有如下要点：

(1) 特征函数对一切分布函数存在，分析性质特别良好．

(2) 特征函数与分布函数一一对应．在绝对可积的假定下特征函数与连续型密度函数构成傅里叶变换对．

(3)《概率论基础》第四章 §5 性质 3 指出了特征函数的一个根本性质即非负定性 (将在下章讲清楚)．

(4)《概率论基础》第四章 §5 性质 4 正如前面所述，预示着特征函数在独立和的研究中起重大作用．

(5) 利用特征函数可方便求取各阶矩．

(6) 记住重要分布的特征函数．

(7) 推广到多元场合，有些异同，为下节做准备．

至此，把特征函数作为处理有限问题工具的基础性工作业已完成．在本节，牛刀小试，用于处理若干重要分布的再生性；下节用

于处理多元正态分布.

为用于处理极限定理, 尚需研究收敛性概念以及建立特征函数与分布函数对应的某种连续性, 将在下章进行.

§ 4.6 多元正态分布

本节暂离叙述主线去完成一项十分重要的任务——研究多元正态分布.

正态分布是概率论中最重要的分布, 而多元正态分布在多元分布中所占的份额还要大得多. "不懂多元正态分布就不能说懂得正态分布"——相信读过本节之后, 你会信服这种讲法.

不懂多元正态分布, 将来在数理统计与随机过程的学习中会处处遇见麻烦, 笔者学生时代就遇见这个问题, 因此有机会写教科书时就觉得非下功夫把这部分写好不可.

二元场合在第三章已作推演, 要推广到多元场合还要有新的工具. 不外两条途径, 一用矩阵论走代数的路, 相对初等; 另一用特征函数走分析的路, 比较艰深, 也处理得比较彻底. 我们是为研究极限定理而引进特征函数并对它作详尽讨论的, 并非专为多元正态分布而来, 但到了这里当然就顺理成章用它作为工具来研究多元正态分布.

用多元特征函数研究多元正态分布还有一个好处就是能处理退化的场合, 在数理统计和随机过程中遇到的多元正态分布, 退化场合并非罕见, 所以更有研究之必要.

因此本节的特点是用特征函数而非密度函数定义 n 元正态分布. 国内后来出的教材也大都沿用这种讲法.

之后的讨论沿着如下线索:

(1) 边际分布仍为正态;

(2) 数字特征;

(3) 不相关等价于独立;

(4) 线性变换下保持正态性不变. 其中《概率论基础》中定理

4.6.6 多维正态变量的线性组合为正态变量的结论提供了定义与研究多元正态分布的一种新途径. 另外, 这段的有些讨论为统计学作了准备.

(5) 把典型分解即《概率论基础》中 (3.2.23) 推广到多元场合, 为回归分析作重要准备.

多元正态分布是概率论与数理统计中最重要的多元分布, 至少表现在以下三个方面:

第一, 它的常见性. 多元正态分布在多元数据 (多元统计数据, 时间序列数据与随机过程数据) 中最为常见, 即使原始数据非正态, 其平均值也很接近于正态 (理由见下章要讲的中心极限定理).

第二, 它具有良好的性质. 多元正态分布性质小结如下:

(i) 多元正态分布由均值向量与协方差阵完全决定.

(ii) 多元正态分布的边际分布仍为正态分布.

(iii) 多元正态分布的条件分布仍为正态分布, 条件均值是线性函数, 条件方差只与联合分布方差阵有关.

(iv) 多元正态变量不相关等价于独立.

(v) 多元正态变量在线性变换与线性组合中保持正态性不变.

(vi) 多元正态变量可经正交变换化为独立正态变量.

第三, 许多多元分布, 例如多项分布, 多元超几何分布, 以它为极限. 不少多元分布由它导出, 例如对应于 χ^2 的 Wishart 分布和对应于 t^2 的 Hotelling T^2 分布.

习题解答与评注

1. (1) 证明关于示性函数成立如下公式:

$$\mathbf{1}_{\overline{A}} = 1 - \mathbf{1}_A;\ \ \mathbf{1}_{AB} = \mathbf{1}_A \cdot \mathbf{1}_B;\ \ \mathbf{1}_{A\cup B} = 1 - \mathbf{1}_{\overline{A}\,\overline{B}}$$

(2) 利用示性函数导出概率加法公式 (1.5.6).

(3) 证明: $\mathbf{1}_{A\cup B\cup C} = \mathbf{1}_A + \mathbf{1}_B + \mathbf{1}_C - \mathbf{1}_{AB} - \mathbf{1}_{BC} - \mathbf{1}_{CA} + \mathbf{1}_{ABC}$

(4) 证明: $\mathbf{1}_{ABC} = \mathbf{1}_A + \mathbf{1}_B + \mathbf{1}_C - \mathbf{1}_{A\cup B} - \mathbf{1}_{B\cup C} - \mathbf{1}_{C\cup A} + \mathbf{1}_{A\cup B\cup C}$

(5) 有五个队参加的比赛中, 每个队与别的队都比赛一场, 若每场比赛参赛双方各有 50% 赢的机会, 试求整个比赛既没有不败的队也没有不胜的队的概率.

(6) 证明 A 与 B 独立的充要条件为 $\mathbf{1}_A$ 与 $\mathbf{1}_B$ 独立.

解与证 (1) **提示:** 用定义, 作验证.

(2)
$$\begin{aligned}\mathbf{1}_{A\cup B} &= 1-\mathbf{1}_{\overline{A}\,\overline{B}}\\ &= 1-\mathbf{1}_{\overline{A}}\cdot\mathbf{1}_{\overline{B}}\\ &= 1-(1-\mathbf{1}_A)(1-\mathbf{1}_B)\\ &= \mathbf{1}_A+\mathbf{1}_B-\mathbf{1}_{AB}\end{aligned}$$

$$E\mathbf{1}_{A\cup B} = E\mathbf{1}_A+E\mathbf{1}_B-E\mathbf{1}_{AB}$$

所以

$$P(A\cup B)=P(A)+P(B)-P(AB)$$

(3)
$$\begin{aligned}\mathbf{1}_{A\cup B\cup C} &= \mathbf{1}_A+\mathbf{1}_{B\cup C}-\mathbf{1}_{AB\cup AC}\\ &= \mathbf{1}_A+\mathbf{1}_B+\mathbf{1}_C-\mathbf{1}_{BC}-\mathbf{1}_{AB}-\mathbf{1}_{AC}+\mathbf{1}_{ABC}\end{aligned}$$

(4)
$$\begin{aligned}\mathbf{1}_{ABC} &= 1-\mathbf{1}_{\overline{A}\cup\overline{B}\cup\overline{C}}\\ &= 1-\left(\mathbf{1}_{\overline{A}}+\mathbf{1}_{\overline{B}}+\mathbf{1}_{\overline{C}}-\mathbf{1}_{\overline{A}\,\overline{B}}-\mathbf{1}_{\overline{B}\,\overline{C}}-\mathbf{1}_{\overline{C}\,\overline{A}}+\mathbf{1}_{\overline{A}\,\overline{B}\,\overline{C}}\right)\\ &= 1-\big(1-\mathbf{1}_A+1-\mathbf{1}_B+1-\mathbf{1}_C+\mathbf{1}_{A\cup B}-1\\ &\qquad +\mathbf{1}_{B\cup C}-1+\mathbf{1}_{C\cup A}-1+1-\mathbf{1}_{A\cup B\cup C}\big)\\ &= \mathbf{1}_A+\mathbf{1}_B+\mathbf{1}_C-\mathbf{1}_{A\cup B}-\mathbf{1}_{B\cup C}-\mathbf{1}_{C\cup A}+\mathbf{1}_{A\cup B\cup C}\end{aligned}$$

(5) 记有不败的队的事件为 A, 有不胜的队的事件为 B, 则所求的概率为 $P(\overline{A}\,\overline{B})$, 而

$$\begin{aligned}\mathbf{1}_{\overline{A}\,\overline{B}} &= (1-\mathbf{1}_A)(1-\mathbf{1}_B)\\ &= 1-\mathbf{1}_A-\mathbf{1}_B+\mathbf{1}_{AB}\end{aligned}$$

现以 A_i 记第 i 队在整个比赛中不败, 以 B_i 记第 i 队在整个比赛中不胜, 则

$$A = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5, \quad B = B_1 \cup B_2 \cup B_3 \cup B_4 \cup B_5$$

因两队比赛总有一方胜一方负, 可知当 $i \neq j$ 时, $P(A_iA_j) = 0$, $P(B_iB_j) = 0$, 且有 $P(A_iB_i) = 0$, 于是,

$$P(A) = \sum_{i=1}^{5} P(A_i) - \sum_{i \neq j} P(A_iA_j) + \cdots = 5 \times \frac{1}{2^4}$$

$$P(B) = \sum_{i=1}^{5} P(B_i) - \sum_{i \neq j} P(B_iB_j) + \cdots = 5 \times \frac{1}{2^4}$$

$$P(AB) = \sum_{i \neq j} P(A_iB_j) - \cdots = \binom{5}{1}\binom{4}{1}\frac{1}{2^4}\frac{1}{2^3}$$

故

$$\begin{aligned} P(\overline{A}\,\overline{B}) &= 1 - P(A) - P(B) + P(AB) \\ &= 1 - \frac{5}{2^4} - \frac{5}{2^4} + \binom{5}{1}\binom{4}{1}\frac{1}{2^4}\frac{1}{2^3} \\ &= \frac{17}{32} \end{aligned}$$

(6) 由

$$P\{\mathbf{1}_A = 1, \mathbf{1}_B = 1\} = P(AB)$$

$$P\{\mathbf{1}_A = 1, \mathbf{1}_B = 0\} = P(A\overline{B})$$

$$P\{\mathbf{1}_A = 0, \mathbf{1}_B = 1\} = P(\overline{A}B)$$

$$P\{\mathbf{1}_A = 0, \mathbf{1}_B = 0\} = P(\overline{A}\,\overline{B})$$

及

$$P\{\mathbf{1}_A = 1\} = P(A), \quad P\{\mathbf{1}_A = 0\} = P(\overline{A})$$

$$P\{\mathbf{1}_B = 1\} = P(B), \quad P\{\mathbf{1}_B = 0\} = P(\overline{B})$$

可知 A 与 B 独立的充要条件即是 $\mathbf{1}_A$ 与 $\mathbf{1}_B$ 的独立.

【评注】示性函数一方面沟通了事件与随机变量, 又能把事件运算化为数值运算, 利于推导公式, 另一方面又沟通了概率与数学期望, 含意深刻, 应用良多.

2. 随机变量 μ 取非负整数值 $n \geqslant 0$ 的概率为 $p_n = A\dfrac{B^n}{n!}$, 已知 $E\mu = a$, 试决定 A 与 B.

提示: 定 A, B 使成均值为 a 的概率分布.

答 $A = \mathrm{e}^{-a}$, $B = a$.

【评注】规范化决定一个常数, 均值也决定一个常数.

3. 设随机变量 ξ 只取非负整数值, 其概率为 $P\{\xi = k\} = \dfrac{a^k}{(1+a)^{k+1}}$, $a > 0$ 是常数, 试求 $E\xi$ 及 $D\xi$.

提示: 代公式, 求级数和.

答 $E\xi = a$, $D\xi = a(a+1)$.

【评注】答案按定义经级数求和可得出. 但若觉察本题的 ξ 服从《概率论基础》中提到的另一种 (只记失败次数的) 几何分布 (3.1.22), 参数 $p = \dfrac{1}{1+a}$, 其均值为 $\dfrac{q}{p}$, 方差为 $\dfrac{q}{p^2}$, 则可省去求和运算.

4. 若事件 A 在第 i 次试验中出现的概率为 p_i, 设 μ 是事件 A 在起初 n 次独立试验中的出现次数, 试求 $E\mu$ 及 $D\mu$.

解 记

$$\xi_i = \begin{cases} 1, & \text{第 } i \text{ 次试验出现 } A \\ 0, & \text{第 } i \text{ 次试验出现 } \bar{A} \end{cases}$$

则 $\mu = \xi_1 + \cdots + \xi_n$, 而 $E\xi_i = p_i$, $D\xi_i = p_i(1-p_i)$, 从而

$$E\mu = \sum_{i=1}^{n} p_i, \quad D\mu = \sum_{i=1}^{n} p_i(1-p_i)$$

【评注】把 μ 表示为计数变量 ξ_i 之和, 是这类问题的解题诀

穷. 0-1 变量的均值、方差公式自当牢记.

5. 一袋中含有 a 只白球, b 只黑球, 从中摸出 c 只 $(c \leqslant a+b)$, 求摸出白球数 μ 的数学期望.

解 记

$$\xi_i = \begin{cases} 1, & \text{第 } i \text{ 次摸到白球} \\ 0, & \text{第 } i \text{ 次摸到黑球} \end{cases}$$

而

$$E\xi_i = \frac{a}{a+b}, \quad i = 1, 2, \cdots, c$$

所以

$$E\mu = E\sum_{i=1}^{c} \xi_i = c \cdot \frac{a}{a+b} = \frac{ca}{a+b}$$

【评注】利用数学期望的可加性, 把 μ 作和式分解; 概率计算用到抽签与顺序无关. 用这种办法化解了题中摸球的前后相关性难题. 本题实质是求超几何分布的数学期望.

6. 试求: (1) 为收集 N 张赠券中的 r 张所需购买的食品袋数 ξ 的数学期望; (2) 为集齐水浒 108 将, 平均要购多少袋? (参照习题一 38 题)

解 (1) 将收集齐第 $j-1$ 张赠券之后到收集齐第 j 张赠券时所购买的食品袋数记为 η_j, 易知

$$\xi = \eta_1 + \eta_2 + \cdots + \eta_r$$

且 η_j 服从几何分布

$$P\{\eta_j = k\} = \left(1 - \frac{N-j+1}{N}\right)^{k-1} \frac{N-j+1}{N}, \quad k = 1, 2, \cdots$$

$$E\eta_j = \frac{N}{N-j+1}, \quad j = 1, 2, \cdots, r$$

因此

$$E\xi = E\eta_1 + \cdots + E\eta_r = N\left(\frac{1}{N} + \cdots + \frac{1}{N-r+1}\right)$$

(2) 由 (1) 知, 平均要购

$$108 \times \left(1+\frac{1}{2}+\cdots+\frac{1}{108}\right) \approx 108 \times (\ln 108+0.5772) \approx 568 \text{ 袋.}$$

其中 0.577 2 是欧拉常数的近似值.

【评注】和式分解, 几何分布, 数学期望可加性, 利用这三招解决了这道难题. 再用调和级数近似式, 通过集齐水浒 108 将赠券所需购买袋数的数值计算, 让学生得到一些感性认识. 以上三道题显示了和式分解法在数学期望计算中的强大威力.

7. 试证: 若取非负整数值的随机变量 ξ 的数学期望存在, 则

$$E\xi=\sum_{k=1}^{\infty} P\{\xi \geqslant k\}$$

证

$$\begin{aligned}
\sum_{k=1}^{\infty} P\{\xi \geqslant k\} &= \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} P\{\xi=j\} \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{j} P\{\xi=j\} \\
&= \sum_{j=1}^{\infty} j P\{\xi=j\} \\
&= E\xi
\end{aligned}$$

【评注】有趣又有用的一个公式. 预示《概率论基础》(5.4.49).

8. 若随机变量 ξ 的分布函数为 $F(x)$, 试证:

$$E\xi=\int_{0}^{\infty}[1-F(x)] \mathrm{d} x-\int_{-\infty}^{0} F(x) \mathrm{d} x$$

特别地, 若 ξ 取非负值, 则

$$E\xi=\int_{0}^{\infty}[1-F(x)] \mathrm{d} x$$

证 由 $E\xi$ 存在, 得 $\int_{-\infty}^{+\infty}|x|\,\mathrm{d}F(x)$ 收敛. 故对任何 $A>0$,

$$0\leqslant AF(-A)\leqslant\int_{-\infty}^{-A}|x|\,\mathrm{d}F(x)\longrightarrow 0,\quad A\to+\infty$$

即

$$\lim_{x\to-\infty}xF(x)=0$$

同样有

$$\lim_{x\to+\infty}x[1-F(x)]=0$$

于是

$$\begin{aligned}
E\xi&=\int_{-\infty}^{+\infty}x\,\mathrm{d}F(x)\\
&=\int_{-\infty}^{0}x\,\mathrm{d}F(x)+\int_{0}^{+\infty}x\,\mathrm{d}F(x)\\
&=\int_{-\infty}^{0}x\,\mathrm{d}F(x)-\int_{0}^{+\infty}x\,\mathrm{d}[1-F(x)]\\
&=xF(x)\Big|_{-\infty}^{0}-\int_{-\infty}^{0}F(x)\,\mathrm{d}x\\
&\quad -x[1-F(x)]\Big|_{0}^{+\infty}+\int_{0}^{+\infty}[1-F(x)]\,\mathrm{d}x\\
&=\int_{0}^{+\infty}[1-F(x)]\,\mathrm{d}x-\int_{-\infty}^{0}F(x)\,\mathrm{d}x
\end{aligned}$$

【评注】所证等式可看作数学期望的几何解释. 数学期望存在的充要条件是这两个积分同时收敛.

非负场合的表达式可与上题对照.

9. 若随机变量 ξ 服从拉普拉斯分布, 其密度函数为

$$p(x)=\frac{1}{2\lambda}\mathrm{e}^{-|x-\mu|/\lambda},\quad -\infty<x<\infty,\quad \lambda>0$$

试求 $E\xi$ 及 $D\xi$.

提示: 用定义, 算积分.

答 $E\xi=\mu,\ D\xi=2\lambda^2$.

【评注】连续型与离散型各设直接计算题一个.

10. 若分子的速度的分布密度函数由麦克斯韦分布律给出:

$$p(s)=\sqrt{\frac{2}{\pi}}\cdot\frac{s^2}{\sigma^3}\exp\left(-\frac{s^2}{2\sigma^2}\right),\quad s>0$$

其中 $\sigma>0$ 是常数, 试求分子的平均速度和平均动能 (假定分子的质量等于 m).

提示: 用佚名统计学家公式.

答 记分子速度为 ξ, 则平均速度 $E\xi=\dfrac{2\sqrt{2}\sigma}{\sqrt{\pi}}$, 平均动能 $E\left(\dfrac{1}{2}m\xi^2\right)=\dfrac{3}{2}m\sigma^2$.

【评注】在气体分子运动论中是验证理论与实验的重要公式.

11. 某城市共有 N 辆汽车, 车牌号从 1 到 N, 若随机地 (可重复) 记下 n 辆车的车牌号, 其最大号码为 ξ, 求 $E\xi$.

解

$$P\{\xi=k\}=\frac{k^n-(k-1)^n}{N^n},\quad k=1,2,\cdots,N$$

$$E\xi=\sum_{k=1}^{N}k\cdot\frac{k^n-(k-1)^n}{N^n}=N-\sum_{k=1}^{N-1}\frac{k^n}{N^n}$$

当 N 充分大的时候,

$$\sum_{k=1}^{N-1}\frac{k^n}{N^n}=N\sum_{k=0}^{N-1}\frac{1}{N}\frac{k^n}{N^n}\approx N\int_0^1 x^n\,\mathrm{d}x=N\frac{1}{n+1}$$

所以

$$E\xi=N-\sum_{k=1}^{N-1}\frac{k^n}{N^n}\approx\frac{n}{n+1}N$$

【评注】本题出现的概率分布已在《概率论基础》第一章 §5 例 5 算出. 这里所得期望值公式很有用, 而且可以推广到其他顺序统计量. 此外, 读者也可以对不重复场合作平行讨论.

12. 若 $a \leqslant \xi \leqslant b$, 试证: $D\xi \leqslant \frac{(b-a)^2}{4}$, 并说明等式在何种情况下成立.

证法一 因 $a \leqslant \xi \leqslant b$, 所以

$$D\xi = E\left[\xi - E\xi\right]^2 \leqslant E\left[\xi - \frac{a+b}{2}\right]^2$$
$$\leqslant E\left[b - \frac{a+b}{2}\right]^2 = \frac{(b-a)^2}{4}$$

方差表征随机变量的离散程度, 所以当 $P\{\xi = a\} = P\{\xi = b\} = \frac{1}{2}$ 时, ξ 的方差达到最大, 此时

$$D\xi = E\xi^2 - \left(E\xi\right)^2 = \frac{a^2}{2} + \frac{b^2}{2} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{4}$$

即等式成立.

证法二 记 $\eta = \frac{\xi - a}{b - a}$, 则 $0 \leqslant \eta \leqslant 1$,

$$D\eta = E\eta^2 - (E\eta)^2 \leqslant E\eta - (E\eta)^2 = E\eta(1 - E\eta) \leqslant \frac{1}{4}$$

从而

$$D\xi = (b-a)^2 D\eta \leqslant \frac{(b-a)^2}{4}$$

等式成立充要条件,

$$E\eta = E\eta^2 = \frac{1}{2}, \quad P\{\eta = 0\} = P\{\eta = 1\} = \frac{1}{2}$$

【评注】证法一用到数学期望的极值性质. 后半题考验读者对方差的涵义的理解.

*13. 若 ξ_1, ξ_2 相互独立, 均服从 $N(\mu, \sigma^2)$, 试证:

$$E\max(\xi_1, \xi_2) = \mu + \frac{\sigma}{\sqrt{\pi}}$$

证 设 $\eta_1 = \frac{\xi_1 - \mu}{\sigma}$, $\eta_2 = \frac{\xi_2 - \mu}{\sigma}$, 则 η_1, η_2 是相互独立的标

准正态变量, 且

$$\max(\xi_1,\xi_2)=\max(\sigma\eta_1+\mu,\sigma\eta_2+\mu)=\sigma\max(\eta_1,\eta_2)+\mu$$

记 $X=\max(\eta_1,\eta_2)$, 其分布函数为

$$F_X(x)=P\{\max(\eta_1,\eta_2)<x\}=P\{\eta_1<x,\eta_2<x\}=\left[\Phi(x)\right]^2$$

得 $X=\max(\eta_1,\eta_2)$ 的密度函数

$$\begin{aligned}p_X(x)&=2\Phi(x)\varphi(x)=\frac{2}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\\&=\frac{1}{\pi}\mathrm{e}^{-\frac{x^2}{2}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\end{aligned}$$

由此得

$$\begin{aligned}E\max(\eta_1,\eta_2)&=\int_{-\infty}^{+\infty}xp_X(x)\,\mathrm{d}x\\&=\frac{1}{\pi}\int_{-\infty}^{+\infty}x\left(\mathrm{e}^{-\frac{x^2}{2}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\right)\mathrm{d}x\\&=-\frac{1}{\pi}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\right)\mathrm{d}\,\mathrm{e}^{-\frac{x^2}{2}}\\&=-\frac{1}{\pi}\left[\mathrm{e}^{-\frac{x^2}{2}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\right]_{-\infty}^{+\infty}+\frac{1}{\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\\&=\frac{1}{\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-x^2}\mathrm{d}x=\frac{1}{\sqrt{\pi}}\end{aligned}$$

从而

$$E\max(\xi_1,\xi_2)=\sigma E\max(\eta_1,\eta_2)+\mu=\mu+\frac{\sigma}{\sqrt{\pi}}$$

【评注】本题要旨是求随机变量函数之一的最大值的数学期望, 上述解法中利用独立性先求最大值的分布, 再对它求数学期望, 相当于直接法. 当然还有其他做法, 例如利用佚名统计学家公式. 亦可参看本章习题 *35.

14. 设 $f(x)$ $(0\leqslant x<\infty)$ 是单调非降函数, 且 $f(x)>0$.

对随机变量 ξ, 若 $Ef(|\xi|)<\infty$, 则对任意 $x>0$, $P\{|\xi|\geqslant x\}\leqslant \frac{1}{f(x)}Ef(|\xi|)$.

证 记 ξ 的分布函数为 $F(x)$,

$$\begin{aligned}P\{|\xi|\geqslant x\}&=\int\limits_{|t|\geqslant x}\mathrm{d}F(t)\\&\leqslant\int\limits_{f(|t|)\geqslant f(x)}\mathrm{d}F(t)\\&\leqslant\frac{1}{f(x)}\int_{-\infty}^{+\infty}f(|t|)\,\mathrm{d}F(t)\\&=\frac{1}{f(x)}Ef(|\xi|)\end{aligned}$$

【评注】本题是切比雪夫型不等式之一, 证明方法也类似.

切比雪夫型不等式证明中, 可如正文一样采用逐级缩小, 也可像本题解一样采用逐级放大, 殊途同归, 实质并无二致.

15. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 为正的独立随机变量, 服从相同分布, 密度函数为 $p(x)$, 试证

$$E\left(\frac{\xi_1+\xi_2+\cdots+\xi_k}{\xi_1+\xi_2+\cdots+\xi_n}\right)=\frac{k}{n}$$

证 因 $\left|\frac{\xi_i}{\xi_1+\xi_2+\cdots+\xi_n}\right|\leqslant 1$, 故 $E\frac{\xi_i}{\xi_1+\xi_2+\cdots+\xi_n}$, $i=1,2,\cdots,n$ 存在.

由对称性

$$E\left(\frac{\xi_1}{\xi_1+\cdots+\xi_n}\right)=E\left(\frac{\xi_2}{\xi_1+\cdots+\xi_n}\right)=\cdots=E\left(\frac{\xi_n}{\xi_1+\cdots+\xi_n}\right)$$

又 $E\left(\frac{\xi_1+\xi_2+\cdots+\xi_n}{\xi_1+\xi_2+\cdots+\xi_n}\right)=1$, 故 $E\left(\frac{\xi_i}{\xi_1+\xi_2+\cdots+\xi_n}\right)=\frac{1}{n}$, 从而

$$E\left(\frac{\xi_1+\xi_2+\cdots+\xi_k}{\xi_1+\xi_2+\cdots+\xi_n}\right)=\frac{k}{n}$$

【评注】对正的随机变量 $\xi_1, \xi_2, \cdots, \xi_n$, 若满足某种对称性使得 $\dfrac{\xi_1}{\xi_1+\xi_2+\cdots+\xi_n}, \dfrac{\xi_2}{\xi_1+\xi_2+\cdots+\xi_n}, \cdots, \dfrac{\xi_n}{\xi_1+\xi_2+\cdots+\xi_n}$

有相同的分布, 则结论就能成立.

题中给定密度, 是想让基础不太好的学生用积分式证明, 但在概率论中较高境界是直接用数学期望的性质证明结论.

16. 袋中装有 N 只球, 但其中白球数为随机变量, 只知其数学期望为 n, 试证从该袋中摸一球得到白球的概率为 $\dfrac{n}{N}$.

证 记 ξ 为袋中的白球数, A 为摸一球得到白球的事件, 则

$$E\xi = \sum_{k=0}^{N} kP\{\xi = k\} = n$$

$$\begin{aligned} P(A) &= \sum_{k=0}^{N} P\{A|\,\xi = k\}P\{\xi = k\} \\ &= \sum_{k=0}^{N} \frac{k}{N}P\{\xi = k\} \\ &= \frac{1}{N}\sum_{k=0}^{N} kP\{\xi = k\} \\ &= \frac{n}{N} \end{aligned}$$

【评注】以本题为"引理"可以解决一大类古典概型的概率计算, 见以下四题.

17. 甲袋中装有 a 只白球 b 只黑球, 乙袋中装有 α 只白球 β 只黑球, 现从甲袋中摸出 $c\ (c \leqslant a+b)$ 只球放入乙袋中, 求从乙袋中再摸一球而为白球的概率.

解法一 令 ξ 为 c 只球中的白球数, B 为从乙袋中再摸一球为白球的事件, 则 ξ 取值 $0, 1, 2, \cdots, c$, 服从超几何分布, 且 $E\xi =$

$\dfrac{ca}{a+b}$ (参看本章习题 5).

$$
\begin{aligned}
P(B) &= \sum_{k=0}^{c} P\{\xi=k\}P\{B|\xi=k\} \\
&= \sum_{k=0}^{c} P\{\xi=k\}\frac{\alpha+k}{\alpha+\beta+c} \\
&= \frac{\alpha}{\alpha+\beta+c}\sum_{k=0}^{c} P\{\xi=k\} + \frac{1}{\alpha+\beta+c}\sum_{k=1}^{c} kP\{\xi=k\} \\
&= \frac{\alpha}{\alpha+\beta+c} + \frac{1}{\alpha+\beta+c}E\xi \\
&= \frac{1}{\alpha+\beta+c}\left(\alpha+\frac{ca}{a+b}\right)
\end{aligned}
$$

解法二 从乙袋摸球时, 袋中有 $\alpha+\beta+c$ 只球, 其中有 $\alpha+\xi$ 只白球. 用上题结论 (“引理”) 立得

$$
P(B) = \frac{\alpha+\dfrac{ca}{a+b}}{\alpha+\beta+c}
$$

【评注】比较两种解法, 可以看出, “引理” 正是由这类解题过程中提炼出来的.

以后遇到这种问题, 可放心直接写出答案.

*18. 袋中有 a 只白球 b 只黑球, 每次摸出一球后总是放入一只白球, 这样进行了 n 次之后, 再从袋中摸一只球, 求它是白球的概率.

解 记 A_i 为 “进行 i 次后, 再从袋中摸一只球是白球” 的事件, $i=0,1,2,\cdots,n$, 则

$$
\begin{aligned}
P(A_{i+1}) &= P(A_i)P(A_{i+1}|A_i) + P(\overline{A}_i)P(A_{i+1}|\overline{A}_i) \\
&= P(A_i)P(A_i) + [1-P(A_i)]\left[P(A_i)+\frac{1}{a+b}\right] \\
&= \left(1-\frac{1}{a+b}\right)P(A_i) + \frac{1}{a+b}
\end{aligned}
$$

即 $P(A_i)$ 满足差分方程:

$$P(A_{i+1})-\left(1-\frac{1}{a+b}\right)P(A_i)=\frac{1}{a+b}$$

解上述差分方程, 得其通解

$$P(A_n)=C\left(1-\frac{1}{a+b}\right)^n+1,\quad C\text{待定}$$

由 $P(A_0)=\dfrac{a}{a+b}$ 可得 $C=-\dfrac{b}{a+b}$, 所以所求概率为

$$1-\frac{b}{a+b}\left(1-\frac{1}{a+b}\right)^n$$

【评注】本题也可以对袋中的白球的数学期望建立差分方程求解, 其形式与难度大体与这里写出的相当. 顺便提及,《概率论基础》第一章 §3 例 9 是本题特例, 眼下我们所用的工具已不可同日而语.

另外, 对一阶差分方程亦可如习题二 32 题那样, 导出递推公式, 备各题引用.

*19. 甲袋中有 a 只白球 b 只黑球, 乙袋中有 c 只白球 d 只黑球, 从两袋中各摸出一球, 并交换放入另一袋中, 这样做了 n 次之后, 再从甲袋中摸出一球, 求这球是白球的概率.

解 以 X_i, Y_i 分别记第 i 次交换后甲, 乙两袋中的白球数, 易知 $X_i+Y_i=a+c$, $EX_i+EY_i=a+c$. 再设

$$\xi_i=\begin{cases}1, & \text{第 } i \text{ 次从甲袋中摸出一只白球}\\0, & \text{第 } i \text{ 次从甲袋中摸出一只黑球}\end{cases}$$

$$\eta_i=\begin{cases}1, & \text{第 } i \text{ 次从乙袋中摸出一只白球}\\0, & \text{第 } i \text{ 次从乙袋中摸出一只黑球}\end{cases}$$

又有 $X_i=X_{i-1}+\eta_i-\xi_i$, 由 16 题的结论 (即 "引理") 还有

$$E\xi_i=P\{\xi_i=1\}=\frac{EX_{i-1}}{a+b},\quad E\eta_i=P\{\eta_i=1\}=\frac{EY_{i-1}}{c+d}$$

因此

$$EX_i = EX_{i-1} + E\eta_i - E\xi_i$$
$$= \left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)EX_{i-1} + \frac{a+c}{c+d}$$

即 EX_n 满足差分方程

$$EX_n - \left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)EX_{n-1} = \frac{a+c}{c+d}$$

该差分方程的通解为

$$EX_n = C\left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)^n + \frac{(a+c)(a+b)}{a+b+c+d}$$

其中 C 为待定常数, 由初始条件 $EX_0 = a$, 得

$$EX_n = \frac{(a+c)(a+b)}{a+b+c+d} + \frac{ad-bc}{a+b+c+d}\left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)^n$$

因而得交换 n 次后再从甲袋中摸出一球为白球的概率为 $\frac{EX_n}{a+b}$, 即

$$\frac{a+c}{a+b+c+d} + \frac{ad-bc}{(a+b)(a+b+c+d)}\left(1 - \frac{1}{a+b} - \frac{1}{c+d}\right)^n$$

【评注】"引理" 和数学期望概念的使用使本题解题思路变得十分清晰.

本题比习题二 32 题难度也明显提高.

*20. 现有 n 个袋子, 各装有 a 只白球 b 只黑球, 先从第一个袋子中摸出一球, 记下颜色后就把它放入第二个袋子中, 再从第二个袋子中摸出一球, 记下颜色后就把它放入第三个袋子中, 照这样办法依次摸下去, 最后从第 n 个袋子中摸出一球并记下颜色, 若在这 n 次摸球中所摸得的白球的总数为 S_n, 试求 ES_n.

解 令

$$\xi_i = \begin{cases} 1, & \text{从第 } i \text{ 个袋子中摸出白球} \\ 0, & \text{从第 } i \text{ 个袋子中摸出黑球} \end{cases}$$

则

$$P\{\xi_1=1\}=\frac{a}{a+b}$$

由 16 题 (即 "引理") 得到

$$P\{\xi_2=1\}=\frac{a+\dfrac{a}{a+b}}{a+b+1}=\frac{a}{a+b}$$

由此类推得

$$P\{\xi_i=1\}=\frac{a}{a+b},\quad i=1,2,\cdots,n$$

又 $S_n=\xi_1+\cdots+\xi_n$, 所以 $ES_n=\dfrac{na}{a+b}$.

【评注】几个要点: 和式分解与数学期望计算, 数学期望在古典概率计算中的应用 (即 "引理"), 递推法即差分方程.

21. 在物理实验中, 为测量某物体的重量, 通常要重复测量多次, 最后再把测量记录的平均值作为该物体的重量, 试说明这样做的道理.

答 设 W 为该物体的实际重量, 第 i 次测量的误差为 ε_i, 测量值为 $W+\varepsilon_i$. 通常设 $E\varepsilon_i=0$, $D\varepsilon_i=\sigma^2$, 且各 ε_i 相互独立.

如果测量了 n 次, 测量记录的平均值为

$$\frac{1}{n}\big(W+\varepsilon_1+\cdots+W+\varepsilon_n\big)=W+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i$$

则它的数学期望为

$$E\Big(W+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\Big)=W$$

而方差则为

$$D\Big(W+\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\Big)=\frac{\sigma^2}{n}$$

只及测量一次的 n 分之一, 这就提高了精度.

【评注】算术平均值的使用, 历史久远. 从概率论与随机误差理论的角度探讨使用算术平均值于测量数据分析的优点见于辛普森的文章. 本题的解答显然注重于均值–方差模型的观点.

*22. 若 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立随机变量, $D\xi_i = \sigma_i^2$, 试找 "权" $a_1, a_2, \cdots, a_n$ (它们满足 $\sum_{i=1}^{n} a_i = 1$), 使 $\sum_{i=1}^{n} a_i\xi_i$ 的方差最小.

解 若有某 $\sigma_k = 0$, 则取 $a_k = 1$, 其余 $a_i = 0$, 则此时 $\sum_{i=1}^{n} a_i\xi_i$ 的方差最小.

若所有 $\sigma_i \neq 0$, 则令

$$f(a_1, \cdots, a_n) = D\Big(\sum_{i=1}^{n} a_i\xi_i\Big) = \sum_{i=1}^{n} a_i^2\sigma_i^2$$

$$\Phi(a_1, \cdots, a_n) = \sum_{i=1}^{n} a_i - 1 = 0$$

作拉格朗日函数

$$F(a_1, \cdots, a_n) = \sum_{i=1}^{n} a_i^2\sigma_i^2 + \lambda\Big(\sum_{i=1}^{n} a_i - 1\Big)$$

$$\frac{\partial F}{\partial a_i} = 2a_i\sigma_i^2 + \lambda = 0, \quad i = 1, 2, \cdots, n$$

解得

$$a_i = -\frac{\lambda}{2\sigma_i^2}$$

及

$$\lambda = -\frac{1}{\sum_{k=1}^{n} \frac{1}{2\sigma_k^2}}$$

所以当 $a_i = \dfrac{1}{\sigma_i^2 \sum_{k=1}^{n} \frac{1}{\sigma_k^2}}$ 时, $\sum_{i=1}^{n} a_i\xi_i$ 方差最小.

【评注】答案是选权使它与方差成反比——完全符合直观.

对上题算术平均值的情况作了推广.

23. 甲与乙依下列规则玩随机游戏: 甲从装有 i个 i 号球 $(i=1,2,3,4,5)$ 的袋中随机摸出一球放入密盒中, 让乙猜号. 乙对甲的支付是他猜的号码与真正的号码之差的 (1) 平方; (2) 绝对值. 试对这两种场合, 讨论乙应采取的最佳策略.

解 设甲摸出一球的号码为 ξ, 乙猜的号码为 C, 乙的最佳策略是使平均支付最小, 因此

(1) 当 $C=E\xi$ 时, $E(\xi-C)^2$ 最小, 故可取

$$C=\frac{1^2+2^2+3^2+4^2+5^2}{1+2+3+4+5}=\frac{11}{3}=3\frac{2}{3}$$

所以在每三次游戏中, 一次猜 3, 两次猜 4.

(2) 当 C 为 ξ 的中位数时, $E|\xi-C|$ 最小, 此时应猜

$$C=4$$

【评注】主要用到如下事实: 均值是最小二乘问题之解, 而中位数是最小绝对偏差问题之解.

不妨思考如下问题: 在这两种场合, 甲在每次游戏中应分别给乙多少支付才能使该游戏成为公平博弈.

24. 某海港对停泊船只供给净水, 初始价是每吨 a 元, 以后再供则要加 50% 的附加费; 若用不完造成浪费则每吨加收资源费 $\frac{a}{4}$. 设某轮船的净水用量是密度函数为 $p(x)$ 的随机变量, 为节约其用水总开支, 试求其最佳首次供水量 y.

解 设某船的初始供水量为 y, 用水量为 ξ, 费用为 η, 则有

$$\eta=\begin{cases} ay+\dfrac{a}{4}(y-\xi)\ , & \xi<y \\ ay+\dfrac{3a}{2}(\xi-y), & \xi\geqslant y \end{cases}$$

于是有

$$\begin{aligned}f(y)=E\eta=&\int_0^y\left[ay+\frac{a}{4}y-\frac{a}{4}x\right]p(x)\,\mathrm{d}x\\&+\int_y^{+\infty}\left[ay+\frac{3a}{2}x-\frac{3a}{2}y\right]p(x)\,\mathrm{d}x\\=&\int_0^y\left(\frac{5a}{4}y-\frac{a}{4}x\right)p(x)\,\mathrm{d}x+\int_y^{+\infty}\left(\frac{3a}{2}x-\frac{a}{2}y\right)p(x)\,\mathrm{d}x\\=&\frac{5a}{4}y\int_0^y p(x)\,\mathrm{d}x-\frac{a}{4}\int_0^y xp(x)\,\mathrm{d}x\\&+\frac{3a}{2}\int_y^{+\infty}xp(x)\,\mathrm{d}x-\frac{a}{2}y\int_y^{+\infty}p(x)\,\mathrm{d}x\end{aligned}$$

上式关于 y 求导,

$$\begin{aligned}f'(y)=&\frac{5a}{4}\int_0^y p(x)\,\mathrm{d}x+\frac{5a}{4}yp(y)-\frac{a}{4}yp(y)\\&-\frac{3a}{2}yp(y)-\frac{a}{2}\int_y^{+\infty}p(x)\,\mathrm{d}x+\frac{a}{2}yp(y)\\=&\frac{5a}{4}\int_0^y p(x)\,\mathrm{d}x-\frac{a}{2}\int_y^{+\infty}p(x)\,\mathrm{d}x\\=&\frac{7a}{4}\int_0^y p(x)\,\mathrm{d}x-\frac{a}{2}\end{aligned}$$

令 $f'(y)=0$, 则得最佳供水量 y 应满足 $\int_0^y p(x)\,\mathrm{d}x=\frac{2}{7}$.

【评注】佚名统计学家公式应用之一例, 问题的提法和解法既典型又新颖.

**25. (Black-Scholes 期权定价公式) 若股票价格 S_T 服从对数正态分布, 即 $\ln S_T\sim N\Big(\ln s_t+\Big(r-\frac{\sigma^2}{2}\Big)(T-t),\ \sigma^2(T-t)\Big)$, $t<T$. 试证明该股票的敲定价为 K 的买入期权的价格 $c_t=\mathrm{e}^{-r(T-t)}E\big[\max(S_T-K,0)\big]$ 满足如下 Black-Scholes 公式:

$c_t = s_t\Phi(d_1) - K\mathrm{e}^{-r(T-t)}\Phi(d_2)$, 其中

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left\{\ln\frac{s_t}{K} + \left(r+\frac{\sigma^2}{2}\right)(T-t)\right\},\quad d_2 = d_1 - \sigma\sqrt{T-t}$$

证 因为 S_T 服从对数正态分布, 所以 S_T 可表示为

$$S_T = s_t\exp\left\{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{T-t}\,Z\right\}$$

其中 $Z\sim N(0,1)$.

$$\begin{aligned}c_t &= \mathrm{e}^{-r(T-t)}E\left[\max\left(S_T-K,0\right)\right]\\ &= \mathrm{e}^{-r(T-t)}\\ &\quad\cdot E\left[\max\left(s_t\exp\left\{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{T-t}\,Z\right\}-K,0\right)\right]\\ &= \mathrm{e}^{-r(T-t)}\\ &\quad\cdot\int_{-\infty}^{+\infty}\max\left(s_t\exp\left\{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{T-t}\,z\right\}-K,0\right)\\ &\quad\cdot\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z\end{aligned}$$

只需讨论

$$s_t\exp\left\{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{T-t}\,z\right\}\geqslant K$$

应有

$$\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{T-t}\,z\geqslant\ln\frac{K}{s_t}$$

故

$$z\geqslant\frac{1}{\sigma\sqrt{T-t}}\left\{\ln\frac{K}{s_t}-\left(r-\frac{\sigma^2}{2}\right)(T-t)\right\}$$

记

$$z^* = \frac{1}{\sigma\sqrt{T-t}}\left\{\ln\frac{K}{s_t}-\left(r-\frac{\sigma^2}{2}\right)(T-t)\right\}$$

从而

$$c_t = \mathrm{e}^{-r(T-t)} \int_{z^*}^{+\infty} \left(s_t \exp\left\{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma\sqrt{T-t}\,z\right\}-K\right)$$

$$\cdot \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\,\mathrm{d}z$$

$$= s_t\int_{z^*}^{+\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(z-\sigma\sqrt{T-t})^2}\,\mathrm{d}z - K\,\mathrm{e}^{-r(T-t)}\int_{z^*}^{+\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\,\mathrm{d}z$$

$$= s_t\left[1-\Phi\left(z^*-\sigma\sqrt{T-t}\right)\right]-K\,\mathrm{e}^{-r(T-t)}\left[1-\Phi(z^*)\right]$$

$$= s_t\Phi\left(\sigma\sqrt{T-t}-z^*\right)-K\,\mathrm{e}^{-r(T-t)}\Phi(-z^*)$$

$$= s_t\Phi\left(\frac{1}{\sigma\sqrt{T-t}}\left\{-\ln\frac{K}{s_t}+\left(r+\frac{\sigma^2}{2}\right)(T-t)\right\}\right)$$

$$-K\,\mathrm{e}^{-r(T-t)}\Phi(-z^*)$$

$$= s_t\Phi(d_1)-K\,\mathrm{e}^{-r(T-t)}\Phi(d_2)$$

【评注】当代数理金融学中最有名的结果. 参看教学札记之二十二“数理金融学和概率论”.

26. **帕累托** (Pareto) **分布**的密度函数为

$$p(x)=\begin{cases} rA^r\dfrac{1}{x^{r+1}}, & x \geqslant A \\ 0, & x < A \end{cases}$$

这里 $r>0$, $A>0$. 试指出这分布具有 p 阶矩, 当且仅当 $p<r$.

解 设 ξ 服从帕累托分布, 当 $p<r$ 时,

$$E\xi^p=\int_A^{+\infty} rA^r\frac{x^p}{x^{r+1}}\,\mathrm{d}x = rA^r\int_A^{+\infty}\frac{\mathrm{d}x}{x^{r-p+1}}=\frac{rA^p}{r-p}$$

且由微积分知识容易知道, 当 $p \geqslant r$ 时, 反常积分 $rA^r\displaystyle\int_A^{+\infty}\frac{\mathrm{d}x}{x^{r-p+1}}$ 发散.

【评注】只存在有限阶矩的分布的例子.

t 分布是另一个典型例子.

27. 若 ξ 的密度函数为

$$p(x)=\begin{cases}\dfrac{1}{2|x|(\ln|x|)^2}, & |x|>\mathrm{e}\\ 0, & \text{其他}\end{cases}$$

试证对于任何 $\alpha>0$, $E|\xi|^\alpha=\infty$.

证

$$E|\xi|^\alpha=\int_{\mathrm{e}}^{+\infty}\frac{x^\alpha\mathrm{d}x}{x(\ln x)^2}=\int_{\mathrm{e}}^{+\infty}\frac{\mathrm{d}x}{x^{1-\alpha}(\ln x)^2}$$

因

$$\lim_{x\to+\infty}\frac{x}{x^{1-\alpha}(\ln x)^2}=\lim_{x\to+\infty}\frac{x^\alpha}{(\ln x)^2}=\infty$$

所以 $E|\xi|^\alpha=\displaystyle\int_{\mathrm{e}}^{+\infty}\frac{\mathrm{d}x}{x^{1-\alpha}(\ln x)^2}$ 发散.

【评注】不存在任何矩的分布的例子.

以柯西分布为代表的矩不存在之例, 皆因"肥尾"之故, 在金融学的近代研究中渐受重视, 因它们带来多次金融危机.

28. 若 ξ 服从 $N(\mu,\sigma^2)$, 试求 $E|\xi-\mu|^k$, 其中 k 为正整数.

解

$$E|\xi-\mu|^k=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}|x-\mu|^k\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x$$

令 $u=\dfrac{x-\mu}{\sigma}$, 则

$$\begin{aligned}E|\xi-\mu|^k&=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\sigma^k|u|^k\mathrm{e}^{-\frac{u^2}{2}}\sigma\,\mathrm{d}u\\&=\sqrt{\frac{2}{\pi}}\sigma^k\int_0^{+\infty}u^k\mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u\\&=\sqrt{\frac{2}{\pi}}\sigma^k(k-1)\int_0^{+\infty}u^{k-2}\mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u\end{aligned}$$

于是有

$$E|\xi-\mu|^k=\begin{cases}\sigma^k(k-1)(k-3)\cdots5\cdot3\cdot1, & k\text{ 是偶数}\\ \sigma^k(k-1)(k-3)\cdots6\cdot4\cdot2\sqrt{\dfrac{2}{\pi}}, & k\text{ 是奇数}\end{cases}$$

【评注】很好的练习, 重要的结果. 正态分布存在任何阶矩.

*29. 记 $a_k = E|\xi|^k$, 若 $a_n < \infty$, 试证 $\sqrt[k]{a_k} \leqslant \sqrt[k+1]{a_{k+1}}$, $k = 1, 2, \cdots, n-1$.

证 由柯西–施瓦茨不等式,

$$\left|E|\xi|^{\frac{k-1}{2}}|\xi|^{\frac{k+1}{2}}\right|^2 \leqslant E|\xi|^{k-1}E|\xi|^{k+1}$$

即

$$a_k^2 \leqslant a_{k-1}a_{k+1}$$

从而

$$a_k^{2k} \leqslant a_{k-1}^k a_{k+1}^k$$

取 $k = 1, 2, \cdots$, 且有 $a_0 = 1$, 于是有

$$\begin{aligned}
&a_1^2 \leqslant a_2^1 \\
&a_2^4 \leqslant a_1^2 a_3^2 \\
&a_3^6 \leqslant a_2^3 a_4^3 \\
&\cdots\cdots\cdots\cdots \\
&a_k^{2k} \leqslant a_{k-1}^k a_{k+1}^k
\end{aligned}$$

将上述不等式两边连乘,

$$\begin{aligned}
a_1^2 a_2^4 a_3^6 \cdots a_{k-1}^{2k-2} a_k^{2k} &\leqslant a_2^1 a_1^2 a_3^2 a_2^3 a_4^3 \cdots a_{k-2}^{k-1} a_k^{k-1} a_{k-1}^k a_{k+1}^k \\
&= a_1^2 a_2^4 a_3^6 \cdots a_{k-1}^{2k-2} a_k^{k-1} a_{k+1}^k
\end{aligned}$$

则有

$$a_k^{2k} \leqslant a_k^{k-1} a_{k+1}^k$$

$$a_k^{k+1} \leqslant a_{k+1}^k$$

所以得 $\sqrt[k]{a_k} \leqslant \sqrt[k+1]{a_{k+1}}$, $k = 1, 2, \cdots, n-1$.

【评注】李雅普诺夫定理: $\sqrt[p]{E|\xi|^p}$ 作为 p 的函数是单调不减的. 可推知: 若高阶矩存在, 则低阶矩必存在.

30. 设随机变量 $\xi_1, \xi_2, \cdots, \xi_{m+n}$ $(n > m)$ 是独立的, 有相同的分布并且有有限的方差, 试求 $S = \xi_1 + \cdots + \xi_n$ 与 $T = \xi_{m+1} +$

$\xi_{m+2}+\cdots+\xi_{m+n}$ 两和之间的相关系数.

提示: 寻求公共点.

答 $\dfrac{n-m}{n}$.

【评注】注意到

$$S=\xi_1+\cdots+\xi_m+\xi_{m+1}+\cdots+\xi_n$$

$$T=\xi_{m+1}+\cdots+\xi_n+\cdots+\xi_{m+n}$$

的公共部分为 $\xi_{m+1}+\cdots+\xi_n$.

31. 若 (ξ,η) 的密度函数为

$$p(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\leqslant 1\\ 0, & x^2+y^2>1\end{cases}$$

试验证: ξ 与 η 不相关, 但它们不独立.

证

$$E\xi=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xp(x,y)\,\mathrm{d}x\mathrm{d}y=\int_{-1}^{1}x\,\mathrm{d}x\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}\,\mathrm{d}y=0$$

同理, $E\eta=0$.

$$\mathrm{cov}(\xi,\eta)=E\xi\eta-E\xi E\eta=\int_{-1}^{1}x\,\mathrm{d}x\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}y\frac{1}{\pi}\,\mathrm{d}y=0$$

即 ξ 与 η 不相关. 然而

$$p_\xi(x)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-x^2}, & |x|\leqslant 1\\ 0 \quad , & |x|>1\end{cases}$$

$$p_\eta(y)=\begin{cases}\dfrac{2}{\pi}\sqrt{1-y^2}, & |y|\leqslant 1\\ 0 \quad , & |y|>1\end{cases}$$

即当 $|x|\leqslant 1$, 且 $|y|\leqslant 1$ 时, $p(x,y)\neq p_\xi(x)p_\eta(y)$, 所以 ξ 与 η 不独立.

【评注】由不相关性不能推出独立性的常用例证之一.

32. 某人写好 n 封信, 又写好 n 只信封, 在黑暗中把每封信随意放入某一信封中, 试求放对的信封数 μ 的数学期望及方差.

解 设

$$\xi_i=\begin{cases}1, & \text{第 } i \text{ 封信放入了相对应的信封,}\\ 0, & \text{第 } i \text{ 封信未放入相对应的信封,}\end{cases}\qquad i=1,2,\cdots,n$$

则

$$\mu=\xi_1+\xi_2+\cdots+\xi_n$$

可以算得 $P\{\xi_i=1\}=\dfrac{1}{n}$, 由此得 $E\xi_i=\dfrac{1}{n}$, $E\xi_i^2=\dfrac{1}{n}$. 又可以算得

$$E\xi_i\xi_j=P\{\xi_i\xi_j=1\}=\frac{1}{n}\,\frac{1}{n-1}$$

$$D\xi_i=E\xi_i^2-\left(E\xi_i\right)^2=\frac{1}{n}-\left(\frac{1}{n}\right)^2=\frac{1}{n}\left(1-\frac{1}{n}\right)$$

$$\operatorname{cov}(\xi_i,\xi_j)=E\xi_i\xi_j-E\xi_iE\xi_j=\frac{1}{n(n-1)}-\frac{1}{n^2}=\frac{1}{n^2(n-1)}$$

于是就有

$$E\mu=E\xi_1+\cdots+E\xi_n=n\times\frac{1}{n}=1$$

和

$$\begin{aligned}D\mu&=D\left(\xi_1+\cdots+\xi_n\right)\\&=\sum_{i=1}^{n}D\xi_i+2\sum_{1\leqslant i<j\leqslant n}\operatorname{cov}(\xi_i,\xi_j)\\&=n\times\frac{1}{n}\left(1-\frac{1}{n}\right)+2\binom{n}{2}\frac{1}{n^2(n-1)}\\&=1\end{aligned}$$

【评注】匹配问题中的匹配数, 其数学期望与方差均为 1, 与 n 无关! 习题一 *35 题中 $P_{[k]}$ 以泊松分布 $P(1)$ 为极限, 所以这个结果实非偶然.

33. 设随机变量 $\zeta \sim N(0,1)$, 记 $A = \{\,|\zeta| > 1\,\}$, $B = \{\,|\zeta| > 2\,\}$, 试求 $\mathbf{1}_A$ 及 $\mathbf{1}_B$ 的概率分布列联表, 数学期望, 方差, 它们的相关系数以及 $D(\mathbf{1}_A + \mathbf{1}_B)$.

答

$\mathbf{1}_B$ \ $\mathbf{1}_A$	0	1	
0	0.682 7	0.271 8	0.954 5
1	0	0.045 5	0.045 5
	0.682 7	0.317 3	

$$E\mathbf{1}_A = 0.317\,3, \quad D\mathbf{1}_A = 0.216\,6;$$

$$E\mathbf{1}_B = 0.045\,5, \quad D\mathbf{1}_B = 0.043\,4;$$

$$\rho_{\mathbf{1}_A\mathbf{1}_B} = \frac{0.045\,5 - 0.317\,3 \times 0.045\,5}{\sqrt{0.216\,6}\sqrt{0.043\,4}} = 0.320\,4;$$

$$D(\mathbf{1}_A + \mathbf{1}_B) = 0.322\,1.$$

【评注】解开一道浅显数值题, 重温几多重要老概念.

*34. 若 ξ, η 服从二元正态分布, $E\xi = a$, $D\xi = 1$, $E\eta = b$, $D\eta = 1$, 证明: ξ 与 η 的相关系数 $r = \cos q\pi$, 其中 $q = P\{(\xi - a)(\eta - b) < 0\}$.

证法一 由题设得

$$q = \frac{1}{2\pi\sqrt{1-r^2}} \iint\limits_{(x-a)(y-b)<0} \mathrm{e}^{-\frac{1}{2(1-r^2)}\left[(x-a)^2 - 2r(x-a)(y-b) + (y-b)^2\right]}\,\mathrm{d}x\mathrm{d}y$$

令 $x = a + \rho\cos\theta$, $y = b + \rho\sin\theta$, $|J| = \rho$,

$$q = \frac{1}{\pi\sqrt{1-r^2}} \int_{\frac{\pi}{2}}^{\pi} \mathrm{d}\theta \int_0^{+\infty} \rho\,\mathrm{e}^{-\frac{\rho^2}{2(1-r^2)}(1-2r\sin\theta\cos\theta)}\,\mathrm{d}\rho$$

$$
\begin{aligned}
&= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi}\frac{\mathrm{d}\theta}{1-2r\sin\theta\cos\theta}\\
&= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi}\frac{\mathrm{d}\theta}{\cos^2\theta(\sec^2\theta-2r\tan\theta)}\\
&= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi}\frac{\mathrm{d}\tan\theta}{1+\tan^2\theta-2r\tan\theta}\\
&= \frac{\sqrt{1-r^2}}{\pi}\int_{\frac{\pi}{2}}^{\pi}\frac{\mathrm{d}(\tan\theta-r)}{1-r^2+(\tan\theta-r)^2}\\
&= \frac{\sqrt{1-r^2}}{\pi}\left(\frac{1}{\sqrt{1-r^2}}\arctan\frac{-r}{\sqrt{1-r^2}}+\frac{\pi}{2\sqrt{1-r^2}}\right)\\
&= \frac{1}{\pi}\arctan\frac{-r}{\sqrt{1-r^2}}+\frac{1}{2}
\end{aligned}
$$

于是得

$$
\tan\left(q\pi-\frac{\pi}{2}\right)=-\frac{r}{\sqrt{1-r^2}}
$$

$$
\cot q\pi=\frac{r}{\sqrt{1-r^2}}
$$

由此推得 $r=\cos q\pi$.

证法二

$P\{(\xi-a)(\eta-b)<0\}$

$$
=\frac{1}{2\pi\sqrt{1-r^2}}\iint\limits_{(x-a)(y-b)<0}\mathrm{e}^{-\frac{1}{2(1-r^2)}\left[(x-a)^2-2r(x-a)(y-b)+(y-b)^2\right]}\mathrm{d}x\mathrm{d}y
$$

$$
\xlongequal[v=y-b]{u=x-a}\frac{1}{2\pi\sqrt{1-r^2}}\iint\limits_{uv<0}\mathrm{e}^{-\frac{1}{2(1-r^2)}(u^2-2ruv+v^2)}\mathrm{d}u\mathrm{d}v
$$

令 $x'=\dfrac{u-rv}{\sqrt{1-r^2}}$, $y'=v$, 该线性变换将 (u,v) 空间的直线 $v=0$ 和 $u=0$ 分别映射为 (x',y') 空间中的直线 $y'=0$ 和 $y'=$

$-\dfrac{\sqrt{1-r^2}}{r}x'$, 若记 $\tan q\pi = \dfrac{\sqrt{1-r^2}}{r}$ (其中 $0 \leqslant q \leqslant 1$), 则直线 $y' = -\dfrac{\sqrt{1-r^2}}{r}x' = -\tan q\pi \cdot x' = \tan(\pi - q\pi)\cdot x'$, (u,v) 空间中的区域 $uv < 0$ 在前述的变换下映射为 (x',y') 空间中的区域 $0 \leqslant y' \leqslant \tan(\pi - q\pi)\cdot x'$ 和 $\tan(\pi - q\pi)\cdot x' \leqslant y' \leqslant 0$,

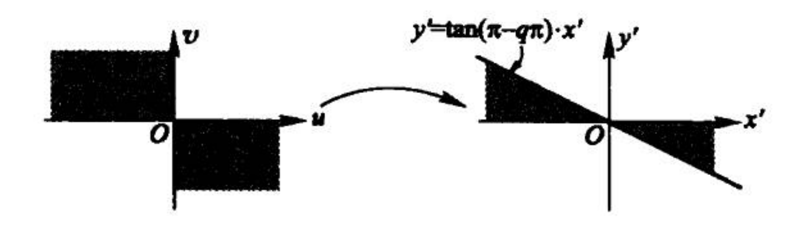


图 4−1

于是延续上面的等式,

$$P\{(\xi-a)(\eta-b)<0\}$$

$$= \frac{1}{2\pi\sqrt{1-r^2}}\iint\limits_{uv<0} e^{-\frac{1}{2(1-r^2)}(u^2-2ruv+v^2)}\,\mathrm{d}u\mathrm{d}v$$

$$\xlongequal[y'=v]{x'=\frac{u-rv}{\sqrt{1-r^2}}} \iint\limits_{\substack{0\leqslant y'\leqslant\tan(\pi-q\pi)\cdot x'\ (x'<0)\\ \tan(\pi-q\pi)\cdot x'\leqslant y'\leqslant 0\ (x'>0)}} \frac{1}{2\pi} e^{-\frac{1}{2}(x'^2+y'^2)}\,\mathrm{d}x'\mathrm{d}y'$$

$$\xlongequal[y'=\rho\sin\theta]{x'=\rho\cos\theta} \frac{1}{2\pi}\int_0^{\infty}\int_{\pi-q\pi}^{\pi} e^{-\frac{\rho^2}{2}}\rho\mathrm{d}\rho\mathrm{d}\theta + \frac{1}{2\pi}\int_0^{\infty}\int_{-q\pi}^{0} e^{-\frac{\rho^2}{2}}\rho\mathrm{d}\rho\mathrm{d}\theta$$

$$= q\int_0^{\infty} e^{-\frac{\rho^2}{2}}\rho\mathrm{d}\rho = q$$

由 $\tan q\pi = \dfrac{\sqrt{1-r^2}}{r}$ 可推得 $r = \cos q\pi$.

【评注】繁复的计算, 简单的结论.

*35. 设 (ξ,η) 服从二元正态分布, $E\xi=E\eta=0$, $D\xi=D\eta=1$, $r_{\xi\eta}=\rho$, 试证

$$E\max(\xi,\eta)=\sqrt{\frac{1-\rho}{\pi}}$$

证

$$\max(\xi,\eta)=\frac{1}{2}(\xi+\eta+|\xi-\eta|)$$

$$E\max(\xi,\eta)=\frac{1}{2}\Big(E\xi+E\eta+E|\xi-\eta|\Big)=\frac{1}{2}E|\xi-\eta|$$

利用多维正态变量的性质可知 $\xi-\eta\sim N\big(0,\ 2(1-\rho)\big)$, 所以

$$E|\xi-\eta|=\frac{1}{\sqrt{2\pi}\sqrt{2(1-\rho)}}\int_{-\infty}^{+\infty}|x|\,\mathrm{e}^{-\frac{x^2}{4(1-\rho)}}\,\mathrm{d}x$$

$$=\frac{2}{\sqrt{2\pi}\sqrt{2(1-\rho)}}\int_{0}^{+\infty}x\,\mathrm{e}^{-\frac{x^2}{4(1-\rho)}}\,\mathrm{d}x$$

$$=-\frac{1}{\sqrt{\pi(1-\rho)}}\,\frac{4(1-\rho)}{2}\,\mathrm{e}^{-\frac{x^2}{4(1-\rho)}}\Big|_0^{+\infty}$$

$$=2\sqrt{\frac{1-\rho}{\pi}}$$

因此有 $E\max(\xi,\eta)=\sqrt{\dfrac{1-\rho}{\pi}}$.

【评注】这里的解法利用正态变量在线性变换下保持正态性不变这一重要性质, 做得干净利落, 比本章习题 *13 的解法也更有普遍性.

36. 甲袋中装有 5 只白球, 7 只黑球, 3 只红球, 乙袋中装有 4 只白球, 4 只黑球, 7 只红球. 试问从哪一个袋中取出一只球有较大不肯定性?

解

$$
\begin{aligned}
H_{甲} &= -\frac{5}{15}\lg\frac{5}{15} - \frac{7}{15}\lg\frac{7}{15} - \frac{3}{15}\lg\frac{3}{15} \\
&= \lg 15 - \frac{1}{15}(5\lg 5 + 7\lg 7 + 3\lg 3) \\
H_{乙} &= -\frac{4}{15}\lg\frac{4}{15} - \frac{4}{15}\lg\frac{4}{15} - \frac{7}{15}\lg\frac{7}{15} \\
&= \lg 15 - \frac{1}{15}(4\lg 4 + 4\lg 4 + 7\lg 7)
\end{aligned}
$$

因为 $x\lg x$ 为下凸函数, 所以 $\frac{1}{2}(5\lg 5 + 3\lg 3) > 4\lg 4$, 则有 $H_{甲} < H_{乙}$, 故乙袋比甲袋有较大的不肯定性.

【评注】练习熵的计算与初步应用.

37. 试求几何分布的熵.

提示: 依定义再用期望值表达式.

答 $-\lg p - \frac{q}{p}\lg q$.

【评注】考虑熵随 p 值的变化情况.

38. 试求二项分布的熵.

提示: 代公式, 求期望.

答 $-\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}\lg\binom{n}{k} - np\lg p - nq\lg q$.

【评注】离散分布算熵仅举二例, 读者可自造他例.

39. 若以 α 及 β 分别记二进位信道的输入及输出, 已知 $P\{\alpha = 1\} = p$, $P\{\alpha = 0\} = 1 - p$, $P\{\beta = 1 \mid \alpha = 1\} = q$, $P\{\beta = 0 \mid \alpha = 1\} = 1 - q$, $P\{\beta = 1 \mid \alpha = 0\} = r$, $P\{\beta = 0 \mid \alpha = 0\} = 1 - r$, 试求输出中含有输入的信息量.

解 即计算 $I(\alpha,\beta)=H(\beta)-H_\alpha(\beta)$. 其中

$$H(\beta)=-\left[(1-p)(1-r)+p(1-q)\right]\lg\left[(1-p)(1-r)+p(1-q)\right]$$
$$-\left[(1-p)r+pq\right]\lg\left[(1-p)r+pq\right]$$

$$H_\alpha(\beta)=-(1-p)\left[(1-r)\lg(1-r)+r\lg r\right]$$
$$-p\left[(1-q)\lg(1-q)+q\lg q\right]$$

所以

$$I(\alpha,\beta)=(1-p)r\lg\frac{r}{(1-p)r+pq}$$
$$+(1-p)(1-r)\lg\frac{1-r}{(1-p)(1-r)+p(1-q)}$$
$$+p(1-q)\lg\frac{1-q}{(1-p)(1-r)+p(1-q)}$$
$$+pq\lg\frac{q}{(1-p)r+pq}$$

【评注】为计算信息量提供一例. 第二章 §1 例 4 已讨论过这类信道.

**40. 在 12 只金属球中, 混有一只假球, 并且不知道它是比真球轻还是重, 用没有砝码的天平来称这些球. (1) 试问至少需要称多少次才能查出这个假球并确定它是比真球轻或重. (2) 给出一种称球方案.

解 (1) 至少需要称 3 次.

(2) 将 12 只球分成三堆, 分别以 $abcd$, $klmn$, $wxyz$ 标识, 以 α_i 表示第 i 次称量, 其上标表示称量的情况.

先将 $abcd$ 和 $klmn$ 放置天平的两边,

(i) $\alpha_1^{abcd=klmn}$ 表示第一次称, 天平平衡, 可推断假球在 $wxyz$ 里,

将 wx 及 ya 放置天平两边,

$\alpha_2^{wx=ya}$, 则可得 z 是假球. 将 a 和 z 放置天平两边,

$\alpha_3^{a>z}$, 则可知 z 比真球轻.

$\alpha_3^{a<z}$, 则可知 z 比真球重.

$\alpha_2^{wx>ya}$, 可推断假球在 wxy 里, 将 w, x 放置天平两边,

$\alpha_3^{w=x}$, 则推断 y 是假球且较轻.

$\alpha_3^{w>x}$, 则推断 w 是假球且较重.

$\alpha_3^{w<x}$, 则推断 x 是假球且较重.

$\alpha_2^{wx<ya}$, 可推断假球在 wxy 里, 将 w, x 放置天平两边,

$\alpha_3^{w=x}$, 则推断 y 是假球且较重.

$\alpha_3^{w>x}$, 则推断 x 是假球且较轻.

$\alpha_3^{w<x}$, 则推断 w 是假球且较轻.

(ii) $\alpha_1^{abcd>klmn}$, 可推断假球在 $abcd$ 或 $klmn$ 里, 将 abk 及 cdl 放置天平两边,

$\alpha_2^{abk=cdl}$, 则可得假球在 mn 中, 将 m 和 n 放置天平两边,

$\alpha_3^{m>n}$, 则可知 n 是假球且比真球轻.

$\alpha_3^{m<n}$, 则可知 m 是假球且比真球轻.

$\alpha_2^{abk>cdl}$, 可推断 cdk 都是真球, ab 之一可能是假球且较重, l 可能是假球且较轻, 将 a, b 放置天平两边,

$\alpha_3^{a=b}$, 则推断 l 是假球且较轻.

$\alpha_3^{a>b}$, 则推断 a 是假球且较重.

$\alpha_3^{a<b}$, 则推断 b 是假球且较重.

$\alpha_2^{abk<cdl}$, 可推断假球在 cdk 里, 将 c, d 放置天平两边,

$\alpha_3^{c=d}$, 则推断 k 是假球且较轻.

$\alpha_3^{c>d}$, 则推断 c 是假球且较重.

$\alpha_3^{c<d}$, 则推断 d 是假球且较重.

(iii) $\alpha_1^{abcd<klmn}$, 此时可类似于 (ii) 情况操作和判断.

【评注】称球问题是一个传播很广的智力游戏题, 解法很多. 它可用信息论的概念和方法解决. 想进一步了解的读者可参看雅格洛姆兄弟合著的《概率与信息》.

41. 试用母函数法求帕斯卡分布的数学期望及方差.

解 帕斯卡分布为

$$P\{\xi=k\}=\binom{k-1}{r-1}p^r q^{k-r}, \qquad k=r,r+1,\cdots$$

其母函数为

$$P(s)=\sum_{k=r}^{\infty}\binom{k-1}{r-1}p^r q^{k-r}s^k$$

令 $m=k-r$,

$$\begin{aligned}P(s)&=p^r\sum_{m=0}^{\infty}\binom{m+r-1}{r-1}q^m s^{m+r}\\&=p^r s^r\sum_{m=0}^{\infty}\binom{m+r-1}{r-1}(qs)^m\\&=p^r s^r\sum_{m=0}^{\infty}\binom{-r}{m}(-qs)^m\\&=\frac{p^r s^r}{(1-qs)^r}\end{aligned}$$

$$P'(1)=\frac{r}{p}$$

$$P''(1)=\frac{r^2+rq}{p^2}-\frac{r}{p}$$

所以

$$E\xi=P'(1)=\frac{r}{p}$$

$$D\xi=P''(1)+P'(1)-\left[P'(1)\right]^2=\frac{rq}{p^2}$$

【评注】以帕斯卡分布为例, 让读者计算一种分布的母函数并用以计算数字特征.

帕斯卡分布的母函数在本章 §4 例 8 中曾由几何分布母函数导出, 此处给出直接推导.

42. 设 ξ 是一个母函数为 $P(s)$ 的随机变量, 试求下列各概率所对应的母函数:

(1) $P\{\xi > n\}$;　　(2) $P\{\xi = 2n\}$.

解　(1) 设 $P\{\xi > n\}$ 对应的母函数为 $F(s)$, 依定义

$$F(s) = \sum_{n=0}^{\infty} P\{\xi > n\} s^n$$

且

$$\begin{aligned}
P(s) &= \sum_{k=0}^{\infty} P\{\xi = k\} s^k \\
&= P\{\xi = 0\} + \sum_{k=1}^{\infty} P\{\xi > k-1\} s^k - \sum_{k=1}^{\infty} P\{\xi > k\} s^k \\
&= P\{\xi = 0\} + s\,F(s) - F(s) + P\{\xi > 0\} \\
&= s\,F(s) - F(s) + P\{\xi \geqslant 0\} \\
&= (s-1)F(s) + 1
\end{aligned}$$

所以

$$F(s) = \frac{1 - P(s)}{1 - s}$$

(2) 设 $P\{\xi = 2n\}$ 对应的母函数为 $G(s)$, 则

$$\begin{aligned}
G(s) &= \sum_{n=0}^{\infty} P\{\xi = 2n\} s^n \\
&= \sum_{n=0}^{\infty} P\{\xi = 2n\} \left(\sqrt{s}\right)^{2n} \\
&= \frac{1}{2}\left[\sum_{n=0}^{\infty} P\{\xi = n\}\left(\sqrt{s}\right)^n + \sum_{n=0}^{\infty} P\{\xi = n\}\left(-\sqrt{s}\right)^n\right] \\
&= \frac{1}{2}\left[P\left(\sqrt{s}\right) + P\left(-\sqrt{s}\right)\right]
\end{aligned}$$

【评注】本题意在训练学生一种推导能力. 其中 $F(s)$ 的公式在某些场合颇为有用.

**43. 在伯努利试验中, 若试验次数 ν 是随机变量, 试证成功的次数与失败的次数这两个随机变量独立的充要条件是 ν 服从泊松分布.

证 以 N_S 记成功的次数, N_F 记失败的次数, $0 < p < 1$ 为成功的概率. 依假定

$$P\{N_S = a \,|\, \nu = a + b\} = \frac{P\{N_S = a,\ \nu = a + b\}}{P\{\nu = a + b\}}$$

$$= \frac{P\{N_S = a,\ N_F = b\}}{P\{\nu = a + b\}}$$

$$= \binom{a+b}{a} p^a q^b$$

即

$$P\{N_S = a,\ N_F = b\} = \binom{a+b}{a} p^a q^b \cdot P\{\nu = a + b\} \qquad (*)$$

充分性: 若 ν 服从泊松分布 $P(\lambda)$, 即

$$P\{\nu = a + b\} = \frac{\lambda^{a+b}}{(a+b)!}\,\mathrm{e}^{-\lambda}$$

代入 (*) 式得

$$P\{N_S = a,\ N_F = b\} = \frac{(a+b)!}{a!\,b!}\,p^a q^b \,\frac{\lambda^{a+b}}{(a+b)!}\,\mathrm{e}^{-\lambda}$$

$$= \frac{(\lambda p)^a}{a!}\,\mathrm{e}^{-\lambda p} \cdot \frac{(\lambda q)^b}{b!}\,\mathrm{e}^{-\lambda q}$$

故 N_S 与 N_F 相互独立.

必要性: 若 N_S 与 N_F 独立, 则

$$P\{N_S = a,\ N_F = b\} = P\{N_S = a\}\,P\{N_F = b\} = f_1(a)g_1(b)$$

由 (*) 式可知

$$P\{\nu = a + b\} \cdot (a+b)! = f_1(a)g_1(b) \cdot a!\,b!\,p^{-a}q^{-b} \triangleq f(a)g(b)$$

因此存在一个函数 h, 使

$$h(a + b) = f(a)g(b) \qquad (**)$$

对一切非负整数 a, b 成立.

由题意, ν 应有正概率取 0 值, 由 $f(0)g(0)=P\{\nu=0\}>0$ 得 $f(0)\neq 0$, $g(0)\neq 0$.

在 (**) 中令 $a=0$, $b=c$ 得 $h(c)=f(0)g(c)$, 又令 $a=c$, $b=0$, 则又得 $h(c)=f(c)g(0)$, 于是有

$$h(c)=f(0)g(c)=f(c)g(0)$$

两边除以 $f(0)g(0)$ 得

$$\frac{h(c)}{f(0)g(0)}=\frac{g(c)}{g(0)}=\frac{f(c)}{f(0)}\triangleq p(c)$$

代入 (**) 得

$$p(a+b)=p(a)p(b)$$

对一切非负整数 a, b 成立. 记 $p(1)=\lambda$, 推得

$$p(k)=\lambda p(k-1)=\cdots=\lambda^{k-1}p(1)=\lambda^k$$

$$h(k)=f(0)g(0)p(k)=f(0)g(0)\lambda^k$$

$$P\{\nu=a+b\}=f(0)g(0)\frac{\lambda^{a+b}}{(a+b)!}$$

由概率的规范化条件, 得

$$f(0)g(0)=\mathrm{e}^{-\lambda}$$

故

$$P\{\nu=a+b\}=\frac{\lambda^{a+b}}{(a+b)!}\mathrm{e}^{-\lambda},\quad a+b=0,1,2,\cdots$$

因此 ν 服从泊松分布.

【评注】这里给出泊松分布的一种刻画, 是直接取自论文的一道难题. 《概率论基础》提供这一模型原型的内容计有: 习题二 43 题, 习题三 17 题, 第四章 §4 例 10.

*44. 设 $\{\xi_k\}$ 是一串独立的整值随机变量序列, 具有相同概率分布, 考虑和 $\eta=\xi_1+\xi_2+\cdots+\xi_\nu$, 其中 ν 是随机变量, 它与 $\{\xi_k\}$

相互独立, 试用 (1) 母函数法; (2) 直接计算证明

$$E\eta = E\nu \cdot E\xi_k, \quad D\eta = E\nu \cdot D\xi_k + D\nu \cdot (E\xi_k)^2$$

证 (1) 设 ξ_k 的母函数为 $F(s)$, ν 的母函数为 $G(s)$, 则 η 的母函数为 $H(s) = G\big(F(s)\big)$.

$$F(1) = \sum_{j=0}^{\infty} P\{\xi_k = j\}\, 1^j = 1$$

$$H'(s) = G'\big(F(s)\big)F'(s)$$

$$H''(s) = G''\big(F(s)\big)\big[F'(s)\big]^2 + G'\big(F(s)\big)F''(s)$$

$$E\eta = H'(1) = G'\big(F(1)\big)F'(1) = G'(1)F'(1) = E\nu \cdot E\xi_k$$

$$\begin{aligned}
D\eta &= H''(1) + H'(1) - \big[H'(1)\big]^2 \\
&= G''(1)\big[F'(1)\big]^2 + G'(1)F''(1) + G'(1)F'(1) - \big[G'(1)F'(1)\big]^2 \\
&= G'(1)\Big[F''(1) + F'(1) - \big[F'(1)\big]^2\Big] \\
&\quad + \big[F'(1)\big]^2\Big[G''(1) + G'(1) - \big[G'(1)\big]^2\Big] \\
&= E\nu \cdot D\xi_k + D\nu \cdot \big(E\xi_k\big)^2
\end{aligned}$$

$$\begin{aligned}
(2)\; P\{\eta = j\} &= \sum_{n=0}^{\infty} P\{\nu = n\}P\{\xi_1 + \cdots + \xi_n = j \,|\, \nu = n\} \\
&= \sum_{n=0}^{\infty} P\{\nu = n\}P\{\xi_1 + \cdots + \xi_n = j\}
\end{aligned}$$

$$\begin{aligned}
E\eta &= \sum_{j=0}^{\infty} jP\{\eta = j\} \\
&= \sum_{j=0}^{\infty} j \sum_{n=0}^{\infty} P\{\nu = n\}P\{\xi_1 + \cdots + \xi_n = j\} \\
&= \sum_{n=0}^{\infty} P\{\nu = n\} \sum_{j=0}^{\infty} jP\{\xi_1 + \cdots + \xi_n = j\}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} P\{\nu = n\}\, E(\xi_1 + \cdots + \xi_n) \\
&= \sum_{n=0}^{\infty} n P\{\nu = n\}\, E\xi_k \\
&= E\nu \cdot E\xi_k
\end{aligned}
$$

$$
\begin{aligned}
E\eta^2 &= \sum_{j=0}^{\infty} j^2 \sum_{n=0}^{\infty} P\{\nu = n\} P\{\xi_1 + \cdots + \xi_n = j\} \\
&= \sum_{n=0}^{\infty} P\{\nu = n\} E(\xi_1 + \cdots + \xi_n)^2 \\
&= \sum_{n=0}^{\infty} P\{\nu = n\} \left[nE\xi_k^2 + n(n-1)(E\xi_k)^2 \right] \\
&= \sum_{n=0}^{\infty} P\{\nu = n\} \left[nD\xi_k + n^2(E\xi_k)^2 \right] \\
&= \sum_{n=0}^{\infty} nP\{\nu = n\} D\xi_k + \sum_{n=0}^{\infty} n^2 P\{\nu = n\} (E\xi_k)^2 \\
&= E\nu \cdot D\xi_k + E\nu^2 \cdot \left(E\xi_k\right)^2
\end{aligned}
$$

$$
\begin{aligned}
D\eta &= E\eta^2 - \left(E\eta\right)^2 \\
&= E\nu \cdot D\xi_k + E\nu^2 \cdot \left(E\xi_k\right)^2 - \left(E\nu \cdot E\xi_k\right)^2 \\
&= E\nu \cdot D\xi_k + D\nu \cdot \left(E\xi_k\right)^2
\end{aligned}
$$

【评注】这是相当基本的两个公式. 第一个很直观, 第二个稍难. 两法相较, 母函数法在处理数字特征中显然方便得多.

45. 某公共汽车站在 $[0, t]$ 中来到的乘客批数 μ 服从参数为 λt 的泊松分布, 而每批来到的乘客数是随机变量, 来 n 个的概率为 p_n, $n = 0, 1, 2, \cdots$, 试求 $[0, t]$ 中来到乘客数 η 的母函数及数学期望.

解 $\eta=\xi_1+\cdots+\xi_\mu$.

这里 ξ_i 表示第 i 批来到的乘客数, 其母函数为

$$F(s)=\sum_{k=0}^{\infty}p_k s^k$$

μ 的母函数为

$$G(s)=\mathrm{e}^{\lambda t(s-1)}$$

故 η 的母函数为

$$H(s)=G\big(F(s)\big)=\mathrm{e}^{\lambda t(F(s)-1)}=\mathrm{e}^{\lambda t\left(\sum\limits_{k=0}^{\infty}p_k s^k-1\right)}$$

$$H'(s)=\mathrm{e}^{\lambda t(F(s)-1)}\cdot\lambda tF'(s)$$

$$E\eta=H'(1)=\lambda tF'(1)=\lambda t\cdot\sum_{k=1}^{\infty}kp_k$$

【评注】复合泊松分布一例. 其数学期望为 $E\mu\cdot E\xi$, 正如上题所述.

本题以实际问题的形式出现, 一些独立性假定蕴含其中.

46. 试用母函数法证明二项分布、泊松分布与帕斯卡分布的再生性.

证 二项分布 $B(n,p)$ 的母函数为 $(q+ps)^n$.

若 $\xi_1\sim B(n_1,p)$, $\xi_2\sim B(n_2,p)$ 且 ξ_1 与 ξ_2 相互独立, 则 $\xi_1+\xi_2$ 的母函数为 $(q+ps)^{n_1}(q+ps)^{n_2}=(q+ps)^{n_1+n_2}$, 即表明 $\xi_1+\xi_2$ 服从 $B(n_1+n_2,\ p)$, 故对相同的 p, 二项分布关于参数 n 具有再生性.

泊松分布 $P(\lambda)$ 的母函数为 $\mathrm{e}^{\lambda(s-1)}$.

若 $\xi_1\sim P(\lambda_1)$, $\xi_2\sim P(\lambda_2)$ 且 ξ_1 与 ξ_2 相互独立, 则 $\xi_1+\xi_2$ 的母函数为 $\mathrm{e}^{\lambda_1(s-1)}\cdot\mathrm{e}^{\lambda_1(s-1)}=\mathrm{e}^{(\lambda_1+\lambda_2)(s-1)}$, 故 $\xi_1+\xi_2\sim P(\lambda_1+\lambda_2)$, 即泊松分布成立再生性.

帕斯卡分布 $f(k;r,p)$ 的母函数为 $\left(\dfrac{ps}{1-qs}\right)^r$.

若 $\xi_1\sim f(k;r_1,p)$, $\xi_2\sim f(k;r_2,p)$ 且 ξ_1 与 ξ_2 相互独立, 则

$\xi_1+\xi_2$ 的母函数为 $\left(\frac{ps}{1-qs}\right)^{r_1}\left(\frac{ps}{1-qs}\right)^{r_2}=\left(\frac{ps}{1-qs}\right)^{r_1+r_2}$, 故有共同 p 的帕斯卡分布关于参数 r 具有再生性.

【评注】用母函数证离散型分布的再生性与用特征函数证一般分布的再生性可作对照.

*47. 若分布函数 $F(x)=1-F(-x+0)$ 成立, 则称它是**对称的**. 试证分布函数对称的充要条件是它的特征函数是实的偶函数.

证 若 ξ 的分布函数为 $F(x)$, 则

$$P\{-\xi<x\}=P\{\xi>-x\}=1-P\{\xi\leqslant -x\}=1-F(-x+0)$$

因此 ξ 具有对称分布的充要条件是 ξ 与 $-\xi$ 有相同的分布函数.

充分性: 若其特征函数 $f(t)$ 是实的偶函数, 即 $f(t)=\overline{f(t)}=f(-t)$, 而 $f(t)=E\mathrm{e}^{\mathrm{i}t\xi}$, $f(-t)=E\mathrm{e}^{\mathrm{i}t(-\xi)}$, 故 ξ 与 $-\xi$ 有相同的特征函数, 从而也有相同的分布函数, 所以有 $F(x)=1-F(-x+0)$.

必要性: 若 $F(x)=1-F(-x+0)$, 则 ξ 与 $-\xi$ 有相同的分布函数, 因而

$$f(t)=E\,\mathrm{e}^{\mathrm{i}t\xi}=E\,\mathrm{e}^{\mathrm{i}t(-\xi)}=E\,\mathrm{e}^{-\mathrm{i}t\xi}=f(-t)=\overline{f(t)}$$

即 ξ 或 $F(x)$ 的特征函数是实的偶函数.

【评注】下列三事实等价: (1) 分布函数对称; (2) ξ 与 $-\xi$ 同分布; (3) 特征函数为实的偶函数.

48. 试求 $[0,1]$ 均匀分布的特征函数.

解

$$\begin{aligned}f(t)&=\int_0^1 \mathrm{e}^{\mathrm{i}tx}\mathrm{d}x\\&=\int_0^1\cos tx\mathrm{d}x+\mathrm{i}\int_0^1\sin tx\mathrm{d}x\\&=\frac{\sin tx}{t}\bigg|_0^1-\mathrm{i}\,\frac{\cos tx}{t}\bigg|_0^1\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sin t}{t} - \mathrm{i}\,\frac{\cos t}{t} + \frac{\mathrm{i}}{t} \\
&= \frac{\cos t + \mathrm{i}\,\sin t - 1}{\mathrm{i}\,t} \\
&= \frac{\mathrm{e}^{\mathrm{i}\,t} - 1}{\mathrm{i}\,t}
\end{aligned}
$$

所以

$$
f(t) = \begin{cases} \dfrac{\mathrm{e}^{\mathrm{i}\,t} - 1}{\mathrm{i}\,t}, & t \neq 0 \\ 1 \quad, & t = 0 \end{cases}
$$

【评注】先求 $U\left[-\frac{1}{2}, \frac{1}{2}\right]$ 特征函数, 再导出 $U[a,b]$ 的特征函数, 同样可以避开复变函数论.

*49. 一般柯西分布的密度函数为 $p(x) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x-\mu)^2}$, $\lambda > 0$. 试证它的特征函数为 $\mathrm{e}^{\mathrm{i}\mu t - \lambda|t|}$, 利用这个结果证明柯西分布的再生性.

证法一 对于柯西分布的特征函数可用下法直接计算,

$$
\begin{aligned}
f(t) &= \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\,tx} \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x-\mu)^2}\, \mathrm{d}x \\
&\overset{v=\frac{x-\mu}{\lambda}}{=\!=\!=} \frac{1}{\pi} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\,t(\lambda v + \mu)} \frac{1}{1+v^2}\, \mathrm{d}v \\
&= \frac{1}{\pi} \mathrm{e}^{\mathrm{i}\,t\mu} \int_{-\infty}^{\infty} \mathrm{e}^{i\lambda\, tv} \frac{1}{1+v^2}\, \mathrm{d}v \\
&= \frac{2}{\pi} \mathrm{e}^{\mathrm{i}\,t\mu} \int_{0}^{\infty} \frac{\cos \lambda\, tv}{1+v^2}\, \mathrm{d}v \\
&= \frac{2}{\pi} \mathrm{e}^{\mathrm{i}\,t\mu} \int_{0}^{\infty} \frac{\cos \lambda\, |t|\, v}{1+v^2}\, \mathrm{d}v
\end{aligned}
$$

问题归结为计算如下拉普拉斯积分:

$$
J = \int_{0}^{\infty} \frac{\cos \beta x}{1+x^2}\, \mathrm{d}x, \qquad \beta = \lambda\,|t| \geqslant 0
$$

先证明

$$\frac{1}{1+x^2}=\int_0^\infty \mathrm{e}^{-xy}\sin y\,\mathrm{d}y,\quad (x>0)$$

这是因为

$$\begin{aligned}
\int_0^\infty \mathrm{e}^{-xy}\sin y\,\mathrm{d}y &= -\int_0^\infty \mathrm{e}^{-xy}\mathrm{d}\cos y\\
&= -\mathrm{e}^{-xy}\cos y\Big|_0^\infty - x\int_0^\infty \mathrm{e}^{-xy}\cos y\,\mathrm{d}y\\
&= 1 - x\int_0^\infty \mathrm{e}^{-xy}\mathrm{d}\sin y\\
&= 1 - x\,\mathrm{e}^{-xy}\sin y\Big|_0^\infty - x^2\int_0^\infty \mathrm{e}^{-xy}\sin y\,\mathrm{d}y\\
&= 1 - x^2\int_0^\infty \mathrm{e}^{-xy}\sin y\,\mathrm{d}y\\
&= \frac{1}{1+x^2}
\end{aligned}$$

此外有另一等式

$$\begin{aligned}
\int_0^\infty \mathrm{e}^{-xy}\cos\beta x\mathrm{d}x &= \frac{1}{\beta}\int_0^\infty \mathrm{e}^{-xy}\mathrm{d}\sin\beta x\quad (y>0)\\
&= \frac{1}{\beta}\mathrm{e}^{-xy}\sin\beta x\Big|_0^\infty + \frac{y}{\beta}\int_0^\infty \mathrm{e}^{-xy}\sin\beta x\,\mathrm{d}x\\
&= -\frac{y}{\beta^2}\int_0^\infty \mathrm{e}^{-xy}\mathrm{d}\cos\beta x\\
&= -\frac{y}{\beta^2}\mathrm{e}^{-xy}\cos\beta x\Big|_0^\infty - \frac{y^2}{\beta^2}\int_0^\infty \mathrm{e}^{-xy}\cos\beta x\,\mathrm{d}x\\
&= \frac{\frac{y}{\beta^2}}{1+\frac{y^2}{\beta^2}} = \frac{y}{\beta^2+y^2}
\end{aligned}$$

于是

$$\begin{aligned}J &= \int_0^\infty \frac{\cos\beta x}{1+x^2}\,\mathrm{d}x \\ &= \int_0^\infty \cos\beta x\mathrm{d}x\int_0^\infty \mathrm{e}^{-xy}\sin y\,\mathrm{d}y \\ &= \int_0^\infty \sin y\,\mathrm{d}y\int_0^\infty \mathrm{e}^{-xy}\cos\beta x\,\mathrm{d}x \\ &= \int_0^\infty \frac{y\sin y}{\beta^2+y^2}\,\mathrm{d}y \\ &\xlongequal{y=\beta x(\beta>0)} \int_0^\infty \frac{x\sin\beta x}{1+x^2}\,\mathrm{d}x\end{aligned}$$

所以有

$$\frac{\mathrm{d}J}{\mathrm{d}\beta} = -J$$

解此微分方程, 得 $J = C\,\mathrm{e}^{-\beta}$, 当 $\beta = 0$, $J = C = \dfrac{\pi}{2}$, 故有

$$J = \frac{\pi}{2}\,\mathrm{e}^{-\beta}$$

回到柯西分布的特征函数, 则有

$$\begin{aligned}f(t) &= \frac{2}{\pi}\,\mathrm{e}^{\mathrm{i}\,t\mu}\int_0^\infty \frac{\cos\lambda\,|t|\,v}{1+v^2}\,\mathrm{d}v \\ &= \frac{2}{\pi}\,\mathrm{e}^{\mathrm{i}\,t\mu}\cdot\frac{\pi}{2}\,\mathrm{e}^{-\lambda\,|t|} \\ &= \mathrm{e}^{\mathrm{i}\mu\,t-\lambda\,|t|}\end{aligned}$$

上述证明过程中涉及交换积分次序, 交换求导和积分的次序, 还有 J 关于参数 β 的连续性, 这些计算在理论上的依据可查阅微积分学中含参变量积分的内容.

最后证明柯西分布的再生性. 设 ξ_1 服从参数为 μ_1, λ_1 的柯西分布, ξ_2 服从参数为 μ_2, λ_2 的柯西分布, 且 ξ_1 与 ξ_2 相互独立, 则 $\eta = \xi_1 + \xi_2$ 的特征函数为

$$f_\eta(t) = \mathrm{e}^{\mathrm{i}\mu_1\,t-\lambda_1\,|t|}\cdot\mathrm{e}^{\mathrm{i}\mu_2\,t-\lambda_2\,|t|} = \mathrm{e}^{\mathrm{i}(\mu_1+\mu_2)\,t-(\lambda_1+\lambda_2)\,|t|}$$

所以 $\eta=\xi_1+\xi_2$ 服从参数为 $\mu_1+\mu_2$, $\lambda_1+\lambda_2$ 的柯西分布.

证法二 记 $f(t)=\mathrm{e}^{\mathrm{i}\mu t-\lambda|t|}$, 易算得

$$\int_{-\infty}^{+\infty}\left|f(t)\right|\mathrm{d}t=\int_{-\infty}^{+\infty}\mathrm{e}^{-\lambda|t|}\mathrm{d}t=\int_{-\infty}^{0}\mathrm{e}^{\lambda t}\mathrm{d}t+\int_{0}^{+\infty}\mathrm{e}^{-\lambda t}\mathrm{d}t=\frac{2}{\lambda}<\infty$$

由《概率论基础》定理 4.5.3, $f(t)$ 对应的密度函数

$$\begin{aligned}p(x)&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-\mathrm{i}tx}\cdot\mathrm{e}^{\mathrm{i}\mu t-\lambda|t|}\mathrm{d}t\\&=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathrm{e}^{-\mathrm{i}(x-\mu)t-\lambda|t|}\mathrm{d}t\\&=\frac{1}{2\pi}\int_{-\infty}^{0}\mathrm{e}^{-\mathrm{i}(x-\mu)t+\lambda t}\mathrm{d}t+\frac{1}{2\pi}\int_{0}^{+\infty}\mathrm{e}^{-\mathrm{i}(x-\mu)t-\lambda t}\mathrm{d}t\\&=\frac{1}{2\pi}\left(\frac{1}{\lambda-\mathrm{i}(x-\mu)}+\frac{1}{\lambda+\mathrm{i}(x-\mu)}\right)\\&=\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2}\end{aligned}$$

再由唯一性定理知, $f(t)$ 是 $p(x)$ 的特征函数.

【评注】利用拉普拉斯积分对柯西分布的特征函数给出了一种纯微积分的推导.

利用傅里叶变换从密度函数求特征函数或从特征函数求密度函数是很有用的方法, 这里给出一个例证.

50. 若随机变量 ξ 服从柯西分布, $\mu=0$, $\lambda=1$, 而 $\eta=\xi$, 试证关于特征函数成立着

$$f_{\xi+\eta}(t)=f_\xi(t)\cdot f_\eta(t)$$

但是 ξ 与 η 并不独立.

证 因 $f_\xi(t)=\mathrm{e}^{-|t|}$, $f_\eta(t)=\mathrm{e}^{-|t|}$, 所以

$$f_{\xi+\eta}(t)=f_{2\xi}(t)=\mathrm{e}^{-|2t|}=\mathrm{e}^{-2|t|}=f_\xi(t)\cdot f_\eta(t)$$

但显然 ξ 与 η 不独立.

【评注】对概率论而言, 特征函数的性质 4, 即独立和的特征函数等于其特征函数之积, 是最关键的性质. 本题提供一个反例说明

其逆不成立.

51. 设 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立且均服从同一柯西分布, 试证: $\frac{1}{n}(\xi_1+\xi_2+\cdots+\xi_n)$ 与 ξ_1 同分布.

证 记 ξ_k 的特征函数为 $f_{\xi_k}(t)=\mathrm{e}^{\mathrm{i}\mu t-\lambda|t|}$, 则

$$
\begin{aligned}
f_{\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k}(t) &= f_{\sum\limits_{k=1}^{n}\xi_k}\left(\frac{t}{n}\right)=\prod_{k=1}^{n} f_{\xi_k}\left(\frac{t}{n}\right)=\left[f_{\xi_1}\left(\frac{t}{n}\right)\right]^n \\
&= \left[\mathrm{e}^{\mathrm{i}\mu\frac{t}{n}-\lambda\left|\frac{t}{n}\right|}\right]^n=\mathrm{e}^{\mathrm{i}\mu t-\lambda|t|}
\end{aligned}
$$

这正是 ξ_1 的特征函数.

【评注】柯西分布诸多“怪性质”之一.

这是泊松于 1827 年在研究观察误差中发现的, 他比柯西在 1853 年找到最小二乘法的反例时发现“柯西分布”早 26 年!

52. 若 $\xi \sim N(\mu, \sigma^2)$, 试用特征函数法求 $E(\xi-\mu)^k$.

解 $\xi-\mu \sim N(0,\sigma^2)$, 其特征函数为 $f(t)=\mathrm{e}^{-\frac{1}{2}\sigma^2t^2}$, 将 $f(t)$ 展开为麦克劳林 (Maclaurin) 级数

$$
\begin{aligned}
f(t) &= \mathrm{e}^{-\frac{1}{2}\sigma^2t^2} \\
&= 1-\frac{1}{2}\sigma^2t^2+\frac{1}{2!}\frac{1}{2^2}\sigma^4t^4+\cdots+\frac{(-1)^n}{n!}\frac{1}{2^n}\sigma^{2n}t^{2n}+\cdots \\
&= 1-\frac{1}{2}\sigma^2t^2+\frac{1}{4!!}\sigma^4t^4+\cdots+\frac{(-1)^n}{(2n)!!}\sigma^{2n}t^{2n}+\cdots
\end{aligned}
$$

所以

$$
E(\xi-\mu)^k=\frac{f^{(k)}(0)}{\mathrm{i}^k}=\begin{cases}\sigma^k(k-1)!!, & k \text{ 是偶数} \\ 0 \quad\quad\quad\quad\;\;, & k \text{ 是奇数}\end{cases}
$$

【评注】参看《概率论基础》(4.2.51) 式, 也可参照习题四 28 题.

该级数也能用来证明正态分布的特征函数或分布函数由其各阶矩所唯一决定.

*53. 求证: 对于任何实值特征函数 $f(t)$, 以下两个不等式成立:

$$1-f(2t)\leqslant 4\big(1-f(t)\big)$$

$$1+f(2t)\geqslant 2\big(f(t)\big)^2$$

证 以 $F(x)$ 记其对应的分布函数, 因为 $f(t)$ 是实值函数, 所以

$$\begin{aligned}1-f(2\,t)&=\int_{-\infty}^{\infty}\big(1-\cos 2\,tx\big)\,\mathrm{d}F(x)\\&=2\int_{-\infty}^{\infty}\big(1-\cos tx\big)\big(1+\cos tx\big)\,\mathrm{d}F(x)\\&\leqslant 4\int_{-\infty}^{\infty}\big(1-\cos tx\big)\,\mathrm{d}F(x)\\&=4\big(1-f(t)\big)\end{aligned}$$

$$\begin{aligned}1+f(2\,t)&=\int_{-\infty}^{\infty}\Big(1+\cos 2\,tx\big)\,\mathrm{d}F(x)\\&=2\int_{-\infty}^{\infty}\cos^2 tx\,\mathrm{d}F(x)\\&\overset{\text{柯西-施瓦兹不等式}}{\geqslant}2\Big(\int_{-\infty}^{\infty}\cos tx\,\mathrm{d}F(x)\Big)^2\\&=2\big(f(t)\big)^2\end{aligned}$$

【评注】两个有用不等式. 与三角函数的倍角公式相连.

*54. 求证: 如果 $f(t)$ 是相应于分布函数 $F(x)$ 的特征函数, 则对于任何 x 值恒成立

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(t)\,\mathrm{e}^{-\mathrm{i}tx}\mathrm{d}t=F(x+0)-F(x-0)$$

证 记

$$\begin{aligned}I&=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(t)\,\mathrm{e}^{-\mathrm{i}\,tx}\mathrm{d}t\\&=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\Big(\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}\,t(y-x)}\mathrm{d}F(y)\Big)\,\mathrm{d}t\end{aligned}$$

由于 $\left|\mathrm{e}^{\mathrm{i}\,t(y-x)}\right|=1$, 二重积分绝对可积, 由富比尼 (Fubini) 定理可交换积分顺序, 得

$$I=\lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty}\left(\int_{-T}^{T}\mathrm{e}^{\mathrm{i}\,t(y-x)}\mathrm{d}t\right)\mathrm{d}F(y)$$

记

$$\begin{aligned}g(T,y)&=\frac{1}{2T}\int_{-T}^{T}\mathrm{e}^{\mathrm{i}\,t(y-x)}\mathrm{d}t\\&=\begin{cases}1, & x=y\\ \dfrac{1}{2T}\displaystyle\int_{0}^{T}2\cos t(y-x)\,\mathrm{d}t=\dfrac{\sin T(y-x)}{T(y-x)}, & x\neq y\end{cases}\end{aligned}$$

由此 $\left|g(T,y)\right|\leqslant 1$, 且 $\lim\limits_{T\to\infty}g(T,y)=\begin{cases}1, & x=y\\0, & x\neq y\end{cases}$, 由控制收敛定理,

$$I=\int_{-\infty}^{\infty}\lim_{T\to\infty}g(T,y)\,\mathrm{d}F(y)=\int_{\{x\}}\mathrm{d}F(y)=F(x+0)-F(x-0)$$

【评注】利用本题结论可由特征函数算出分布函数一切不连续点的跃度.

55. 随机变量 ξ 的特征函数为 $f(t)$, 且它的 n 阶矩存在, 令

$$\chi_k=\frac{1}{\mathrm{i}^k}\left[\frac{\mathrm{d}^k}{\mathrm{d}t^k}\log f(t)\right]_{t=0},\quad k\leqslant n$$

称 χ_k 为随机变量 ξ 的 k **阶半不变量**.

(1) 试证 $\eta=\xi+b$ (b 是常数) 的 k ($k>1$) 阶半不变量等于 χ_k;

(2) 试求出半不变量与原点矩之间的关系式.

解与证 (1) 由 $\eta=\xi+b$ 得 $f_\eta(t)=\mathrm{e}^{\mathrm{i}\,tb}f_\xi(t)$, 故有

$$\log f_\eta(t)=\mathrm{i}\,tb+\log f_\xi(t)$$

$$\mathrm{i}^{-1}\frac{\mathrm{d}\log f_\eta(t)}{\mathrm{d}t}=b+\mathrm{i}^{-1}\frac{\mathrm{d}\log f_\xi(t)}{\mathrm{d}t}$$

$$\frac{\mathrm{d}^k\log f_\eta(t)}{\mathrm{d}t^k}=\frac{\mathrm{d}^k\log f_\xi(t)}{\mathrm{d}t^k},\quad k>1$$

所以 $\xi+b$ 的 $k(k>1)$ 阶半不变量等于 ξ 的 k 阶半不变量 χ_k.

(2) 设 ξ 的特征函数为 $f(t)$, 它与原点矩有等式

$$f^{(k)}(0)=\mathrm{i}^k m_k=\mathrm{i}^k E\xi^k$$

$$f(t)=1+\sum_{k=1}^{\infty}\frac{m_k}{k!}(\mathrm{i}\,t)^k$$

由半不变量的定义, 又有

$$\log f(t)=\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(\mathrm{i}\,t)^k$$

即

$$\begin{aligned}f(t)&=\exp\Big\{\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(\mathrm{i}\,t)^k\Big\}\\&=1+\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(\mathrm{i}\,t)^k+\frac{1}{2!}\Big[\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(\mathrm{i}\,t)^k\Big]^2+\cdots\end{aligned}$$

由 $f(t)$ 的两个展开式得

$$\sum_{k=1}^{\infty}\frac{m_k}{k!}(\mathrm{i}\,t)^k=\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(\mathrm{i}\,t)^k+\frac{1}{2!}\Big[\sum_{k=1}^{\infty}\frac{\chi_k}{k!}(\mathrm{i}\,t)^k\Big]^2+\cdots$$

比较等式两边 t^k 项的系数, 得

$$m_1=\chi_1,\quad m_2=\chi_2+\chi_1^2,\quad m_3=\chi_3+3\chi_1\chi_2+\chi_1^3,\quad\cdots$$

【评注】半不变量是随机变量或分布的另一类数字特征, 它的特点是在平移下不变, 不过这类数字特征目前已很少使用.

56. 若随机向量 (ξ,η) 服从二元正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 试写出:

(1) (ξ,η) 的特征函数;

(2) $a\xi+b\eta$ 的密度函数;

(3) 当 $\xi=x$ 时 η 的条件密度函数, 并讨论 $\rho=-1,\rho=1$ 及 $\rho=0$ 等特殊情况下结果的概率含意.

答 (1) $\mathrm{e}^{\mathrm{i}(\mu_1 t_1+\mu_2 t_2)+\frac{1}{2}(\sigma_1^2 t_1^2+2\rho\sigma_1\sigma_2 t_1 t_2+\sigma_2^2 t_2^2)}$;

(2) $\dfrac{1}{\sqrt{2\pi}\sqrt{a^2\sigma_1^2+b^2\sigma_2^2+2ab\rho\sigma_1\sigma_2}}\mathrm{e}^{-\frac{(x-a\mu_1-b\mu_2)^2}{2(a^2\sigma_1^2+b^2\sigma_2^2+2ab\rho\sigma_1\sigma_2)}}$;

(3) $\dfrac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\mathrm{e}^{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))\right]^2}$,

$\rho=-1$, ξ 与 η 完全负相关, 条件密度函数不存在,

$$\eta=\mu_2-\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

$\rho=1$, ξ 与 η 完全正相关, 条件密度函数不存在,

$$\eta=\mu_2+\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

$\rho=0$, ξ 与 η 不相关, 推得 ξ 与 η 独立, 条件密度化为无条件密度,

$$\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

【评注】关于二元正态的这些基本事实, 应能熟练写出.

57. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立, 均服从 $N(0,1)$, 而

$$\eta_1=\sum_{k=1}^n a_k\xi_k,\quad \eta_2=\sum_{k=1}^n b_k\xi_k$$

试证 η_1 与 η_2 独立的充要条件为 $\sum\limits_{k=1}^n a_kb_k=0$.

证 首先应指出 (η_1,η_2) 作为 $\xi_1,\xi_2,\cdots,\xi_n$ 的线性变换构成二维正态变量. 又有 $E\eta_1=0$, $E\eta_2=0$, 且

$$\begin{aligned}E\eta_1\eta_2&=E\Big(\sum_{k=1}^n a_k\xi_k\Big)\Big(\sum_{k=1}^n b_k\xi_k\Big)\\&=E\Big(\sum_{k=1}^n a_kb_k\xi_k^2\Big)+E\Big(\sum_{j\neq k}a_jb_k\xi_j\xi_k\Big)\\&=\sum_{k=1}^n a_kb_k\end{aligned}$$

所以 η_1 与 η_2 不相关即 $E\eta_1\eta_2 = E\eta_1 E\eta_2 = 0$ 的条件是 $\sum_{k=1}^{n} a_k b_k = 0$. 由二维正态变量不相关与独立的一致性知 η_1 与 η_2 独立的充要条件为 $\sum_{k=1}^{n} a_k b_k = 0$.

【评注】这也是正态变量的一个刻画性性质.

58. 设 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立, 具有相同分布 $N(\mu, \sigma^2)$, 试求 $\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$ 的分布, 并写出它的数学期望及协方差矩阵, 再求 $\overline{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ 的分布密度.

答

$\boldsymbol{\xi}$ 的联合分布密度函数为 $\dfrac{1}{(\sqrt{2\pi}\sigma)^n}\mathrm{e}^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$

$$E\boldsymbol{\xi} = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}; \quad D\boldsymbol{\xi} = \sigma^2 \boldsymbol{I}_n = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

$$\overline{\xi} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$p_{\overline{\xi}}(x) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{n(x-\mu)^2}{2\sigma^2}}$$

【评注】都是必须牢记的最基本结论. 在统计学中更是常用.

59. 若 (ξ, η) 服从 $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 而

$$U = a\xi + b\eta, \quad V = c\xi + d\eta$$

(1) 试求 U 与 V 的数学期望, 方差及相关系数;

(2) 写出 (U,V) 的分布;

(3) 讨论: 何种情况下, (U,V) 退化为一维分布, 何种情况下, U 与 V 独立.

解 (1) $EU=a\mu_1+b\mu_2,\quad DU=a^2\sigma_1^2+b^2\sigma_2^2+2ab\sigma_1\sigma_2\rho$

$$EV=c\mu_1+d\mu_2,\quad DV=c^2\sigma_1^2+d^2\sigma_2^2+2cd\sigma_1\sigma_2\rho$$

$$\operatorname{cov}(U,V)=ac\sigma_1^2+bd\sigma_2^2+(ad+bc)\sigma_1\sigma_2\rho$$

$$\rho_{UV}=\frac{ac\sigma_1^2+bd\sigma_2^2+(ad+bc)\sigma_1\sigma_2\rho}{\sqrt{a^2\sigma_1^2+b^2\sigma_2^2+2ab\sigma_1\sigma_2\rho}\sqrt{c^2\sigma_1^2+d^2\sigma_2^2+2cd\sigma_1\sigma_2\rho}}$$

(2) 因 $\begin{pmatrix}U\\V\end{pmatrix}=\begin{pmatrix}a&b\\c&d\end{pmatrix}\begin{pmatrix}\xi\\\eta\end{pmatrix}$, 故知 (U,V) 服从二元正态分布 $N\left(EU,\,EV,\,DU,\,DV,\,\rho_{UV}\right)$, 其中的参数由 (1) 给出.

(3) 若记 $\boldsymbol{B}=\begin{pmatrix}DU&\operatorname{cov}(U,V)\\\operatorname{cov}(U,V)&DV\end{pmatrix}$ 则

$$\left|\boldsymbol{B}\right|=(ad-bc)^2\sigma_1^2\sigma_2^2-(ad-bc)^2\sigma_1^2\sigma_2^2\rho^2=(ad-bc)^2(1-\rho^2)\sigma_1^2\sigma_2^2$$

因此当 $a,\,b,\,c,\,d$ 不全为零, $ad=bc$ 或 $\rho=\pm1$ 时, (U,V) 退化为一维变量.

而当 $ac\sigma_1^2+(ad+bc)\sigma_1\sigma_2\rho+bd\sigma_2^2=0$ 时, $\rho_{UV}=0,U,V$ 独立.

【评注】本题提示如下要点: (1) 由线性变换下保持正态性不变断定 (U,V) 服从二元正态, 从而把求分布等问题化为前两阶矩的计算而获得巨大方便; (2) $|\boldsymbol{B}|=0$ 表示 (U,V) 退化; (3) $\rho_{UV}=0$ 表明 U 与 V 独立.

**60. (Fisher 引理) 若 $X_1,X_2,\cdots,X_n$ 相互独立, 均服从 $N(\mu,\sigma^2)$, 记

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

试证:

(1) $\overline{X}$ 与S_n^2 相互独立;

(2) $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$;

(3) $\dfrac{nS_n^2}{\sigma^2} \sim \chi_{n-1}^2$.

证 取正交阵

$$A = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3\cdot 2}} & \frac{1}{\sqrt{3\cdot 2}} & \frac{-2}{\sqrt{3\cdot 2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{pmatrix}$$

记 $\boldsymbol{X}^{\mathrm{T}} = (X_1, X_2, \cdots, X_n)$, $\boldsymbol{Y}^{\mathrm{T}} = (Y_1, Y_2, \cdots, Y_n)$, 且令 $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$, 则可知

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n), \qquad \boldsymbol{Y} \sim N(\boldsymbol{A}\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n)$$

其中 $\boldsymbol{\mu}^{\mathrm{T}} = (\mu, \mu, \cdots, \mu)$, $\boldsymbol{I}_n$ 是 n 阶单位阵. 于是

$$Y_1 = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \sim N(\sqrt{n}\mu, \sigma^2)$$

$$Y_k \sim N(0, \sigma^2), \quad k = 2, 3, \cdots, n$$

且 $Y_1, Y_2, Y_3, \cdots, Y_n$ 相互独立. 又有

$$\begin{aligned} nS_n^2 &= \sum_{k=1}^{n} (X_k - \overline{X})^2 = \sum_{k=1}^{n} X_k^2 - n\overline{X}^2 = \sum_{k=1}^{n} Y_k^2 - Y_1^2 \\ &= Y_2^2 + Y_3^2 + \cdots + Y_n^2 \end{aligned}$$

(1)
$$\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k = \frac{1}{\sqrt{n}} Y_1$$

由 Y_1 与 $Y_2, Y_3, \cdots, Y_n$ 独立, 可得 $\overline{X}$ 与 S_n^2 独立;

(2) 由 $Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$, 推得 $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$;

(3) 从 $\frac{nS_n^2}{\sigma^2}=\frac{Y_2^2}{\sigma^2}+\frac{Y_3^2}{\sigma^2}+\cdots+\frac{Y_n^2}{\sigma^2}$ 可知$\frac{nS_n^2}{\sigma^2}\sim\chi^2_{n-1}$.

【评注】这在正态总体统计学 (近代数理统计是以它为出发点建立起来的) 中是最核心的结果.

习题总评

第四章习题以计算数字特征——数学期望, 方差, 矩, 相关系数为主, 但很少配置硬算、死算的题目, 更强调 "活算" ——主要是用有关性质作巧算. 作者希望通过题目的精选和示范性解答能让使用本教材的学生学到更多这方面的技巧, 也加深对这些数字特征的理解.

对特征函数 (还有母函数) 只配置了适量的题目. 它的威力在多元正态场合已见端倪, 对它在极限定理中有力应用还在下一章.

关于正态分布的一些题目都较重要, 有关技巧应很好掌握.

章后小议

本章引进分布的数字特征与特征函数.

数字特征正如其名, 只刻画随机变量或分布函数的部分特征. 从研究分布函数转到研究数字特征看似退一步 (不能得到变量的全面信息只能用其部分信息), 事实上却是别开生面. 因为后者只用一、二个数就能抓住变量的特征, 而前者则要用一个函数. 何况所有有名的分布都只由一二个参数 (它们与数字特征有紧密的关系) 所决定. 这个优点 (所谓参数化) 在从数据到模型 (这是数理统计研究的主题) 的转换中尤为重要.

最主要的数字特征是数学期望, 它表征随机变量取值的重心, 还因为其他数字特征大都是随机变量的某种函数的数学期望. 《概率论基础》从离散型出发, 直接利用斯蒂尔切斯积分, 对连续型直至一般分布定义数学期望, 并利用一组典型的应用实例来阐述数学

期望的直观意义, 再通过佚名统计学家公式讨论数学期望的性质.

佚名统计学家公式是概率论中最常用的公式之一, 而且在使用它时常常感觉不到它的存在, 真做到 "润物细无声". 学完本章时, 值得把全章检查一遍找出使用它的地方.

对方差的直观含意的解释, 对相关系数既深且广的讨论, 显示数学期望与方差连用的威力的应用实例, 成为新版的一个特色.

简言之, 以平均对抗分散是概率统计的要义, 均值表征中心趋势, 方差则度量其误差.

把熵与信息写入概率论教程不多见, 聊备一格.

特征函数在《概率论基础》中得到充分强调, 这里深受苏联教材影响, 但比起欧美教程主要讲矩母函数, 优点明显多于缺点, 因为后者不是对于所有分布都存在, 因此不能真正解决问题. 讲特征函数要用到复变函数, 基础要求稍高, 但经过仔细处理, 书中并未用到多少复变知识.

在展现特征函数的主要应用 (下章用于研究极限定理) 之前, 用它研究了多元正态分布. 因为国内早期概率论教本对多元正态分布只讲到二元为止, 这在后继的随机过程和数理统计中完全不够用, 自王梓坤《概率论基础及其应用》出版后, 情况开始改变. 《概率论基础》不用密度函数而改用特征函数定义多元正态, 从而包含了退化的场合, 得出更一般的结果.

母函数在《概率论基础》中安排作为特征函数的前导, 并强调了对独立和的处理, 这前后两节的讲述较多采用互见的形式.

教学札记之二十

圣彼得堡悖论与期望效用函数

一、圣彼得堡悖论

1703 年俄国沙皇彼得大帝在波罗的海的沼泽地上建立起新都圣彼得堡. 1725 年彼得大帝去世由其妻叶卡婕琳娜继位, 她在位短

短两年内就建立起圣彼得堡科学院, 实现其夫夙愿.

1725 年 10 月 27 日瑞士伯努利家族第二代的尼古拉・伯努利 (Nicolas Bernoulli, 1695—1726) 和丹尼尔・伯努利 (Daniel Bernoulli, 1700—1782) 兄弟作为应邀的外国院士到达圣彼得堡. 次年, 尼古拉提出了一个概率问题, 此问题后以 "圣彼得堡悖论" 著称, 在 18, 19 世纪非常有名. 不幸, 尼古拉到那里才八个月就死去, 其职位由其弟丹尼尔继承. 他继续研究圣彼得堡悖论, 于 1738 年写成 "**风险测度的新理论注释**", 提出自己的解答. 此文由于对现代经济学的重大价值, 200 年之后由 Econometrica 于 1954 年 1 月译成英文在该刊当年第一期重新刊登.

事实上, 这个问题在 1713 年 9 月 9 日他们的堂哥尼古拉给蒙特莫特的私人通信中已经提出, 被收入蒙特莫特的《随笔》.

所谓 "圣彼得堡悖论" 可表述如下:

彼得掷一枚硬币, 到掷出正面为止, 这时该次赌博就算完结了. 如果在第一次掷出正面, 则巴维尔要给彼得一个卢布; 如果恰好在第 2 次掷出正面, 则给 2 个卢布; 如果恰好在第 3 次掷出正面, 则给 4 个卢布. 一般地, 如果在第 n 次掷出正面, 则巴维尔要给彼得 2^{n-1} 个卢布. 问彼得在赌博开始时应付给巴维尔多少卢布才算公平博弈.

这个问题后来受到丹尼尔, 达朗贝尔, 蒲丰, 孔多塞, 拉普拉斯, 泊松, 博雷尔等著名数学家的讨论. 为什么这么多人对它感兴趣呢?

首先是因为彼得在游戏中的所得 X 的期望值是无穷大. 事实上, X 的取值与概率分布为

$$P\{X = 2^{n-1}\} = 2^{-n}, \quad n = 1, 2, \cdots \tag{1}$$

因此

$$EX = \sum_{n=1}^{\infty} 2^{n-1} \cdot 2^{-n} = \infty \tag{2}$$

世界上不存在有无穷财富的人. 彼得当然没有无穷财富, 因此

这个游戏根本玩不起来. 况且为进行一个"公平博弈"(这是自惠更斯《论赌博中的计算》以来深入人心的一个概念), 要先付出无穷财富才能开始, 也是十分荒谬的.

其次, 属于心理学上的证据. 有人作了测试, 找来一批大学生, 问他们为进行这一游戏愿付多少? 大部分只准备付 2 至 3 卢布来玩这种游戏, 个别人愿付到 8 卢布, 但没有人愿付更多!

二、丹尼尔·伯努利的解法

丹尼尔的解法建立在下述概念的基础上: 人们对奖励所关心的是效用 (Utility) 而非货币价值; 而额外的货币增加所得的额外效用随着奖励的货币价值的增加而减少. 因此他作出特别的假定: 货币效用是货币奖励大小的对数函数, 其函数形式为

$$U(X) = b\log\frac{X}{a} \tag{3}$$

这里 $U(X)$ 是货币量 X 的效用, a, b 是正常数.

丹尼尔认为, 在确定圣彼得堡游戏的价值时, 一个人会考虑由奖励所提供的效用而非它的货币量. 因此他乐于为玩游戏所付的货币数量取决于游戏的期望效用, 而不取决于期望货币报酬.

在圣彼得堡游戏中, 期望效用为

$$\begin{aligned}
EU(X) &= \sum_{n=1}^{\infty}\frac{1}{2^n}\,b\log\frac{2^{n-1}}{a}\\
&= \sum_{n=1}^{\infty}\frac{1}{2^n}\,b\big[(n-1)\log 2 - \log a\big]\\
&= b\log 2\sum_{n=1}^{\infty}\frac{n-1}{2^n} - b\log a\sum_{n=1}^{\infty}\frac{1}{2^n}\\
&= b\log 2 - b\log a = b\log\frac{2}{a}
\end{aligned} \tag{4}$$

因此这个期望效用等于 2 卢布货币的效用, 即 $EU(X) = U(2)$, 由此可得: 具有伯努利效用函数所表示的风险偏好的人只愿付 2 卢

布来参加圣彼得堡游戏. 如果有一项确定性的可超过 2 卢布收入的选择, 他将偏好这种确定性的选择而不管圣彼得堡游戏有可能得到无穷的货币报酬.

丹尼尔把自己引进的期望效用也称为**道德期望**, 并作道德解释.

丹尼尔 · 伯努利在他的论文中还介绍了另一位数学家克拉默独立提出的类似的解法. 克拉默也用效用的数学期望来解决悖论, 不过他引进的效用函数是

$$U(X) = \sqrt{X} \tag{5}$$

针对这种效用函数, 圣彼得堡游戏的期望效用为

$$\begin{aligned} EU(X) &= \sum_{n=1}^{\infty} \frac{1}{2^n}\sqrt{2^{n-1}} \\ &= \frac{1}{2-\sqrt{2}} \end{aligned} \tag{6}$$

因此具有平方根效用函数所表示的风险偏好的人为参加圣彼得堡游戏愿支付的最大价格 x_0 可以由方程

$$EU(X) = \sqrt{x_0}$$

得到, 所以

$$x_0 = \left(\frac{1}{2-\sqrt{2}}\right)^2 = 2.914\ (\text{卢布})$$

这时 2.914 卢布的效用是 $U(2.914) = \sqrt{2.914} = \dfrac{1}{2-\sqrt{2}}$, 正是游戏的期望效用.

对于丹尼尔 · 伯努利为解决圣彼得堡悖论所引进的道德期望概念, 在同时代及后代的数学家中引起很多争论, 有人赞同, 有人则强烈反对.

三、对道德期望的不同评价

同时代的达朗贝尔 (1717—1783) 评论道: “我们不知道他的解

是否令人满意, 有些流言蜚语值得数学家重视. ”

蒲丰 (1707 — 1788) 在 1777 年的一篇文章中则提倡与数学期望平等地引入道德期望的概念: “一个守财奴和一个数学家相似, 都只看钱的票值. 一个明智的人, 不在乎它的常规价值, 只看他能从中得到多少效用. 他的观点比守财奴合理, 他的感受比数学家快活. 被一个穷人用来尽法律义务的一撒勒和填充高利贷者口袋的一撒勒, 在守财奴和数学家的眼中具有同样的价值: 前者认为纳捐可取得 ‘等值’ 的乐趣, 后者把它看成 ‘等值’ 的一个单元. 但一个明智的人, 将穷人的一撒勒评价为一枚金币, 而高利贷者的一撒勒看成只值一分钱. ”

贝特朗 (1822 — 1900) 在他的文章中挖苦说: “道德期望成了一门经典理论, 这里的‘经典’是反义的. 这门理论被研究、教授并在相当有名气的书上被描述, 但是它的成功到此为止, 没有实际的应用, 没有做过什么切实可行的事情. ”

格涅坚科 (1912 — 1995) 在《概率论教程》的附录 “概率论简史” 中介绍了彼得堡悖论, 但他说: “彼得堡院士丹尼尔 · 伯努利导入了所谓 ‘道德期望’ 这个概念. 我们不来讲这个概念的定义了, 因为它没有科学的意义, 并且, 我顺便说, 它也不解决彼得堡赌法的奇论. ”

笔者从 60 年代起就相信格涅坚科的这段话, 因为我的概率知识最早就是从他的书中学来的, 直到 90 年代中期学习了西方金融理论才知道这个评价是欠公平的.

四、效用理论在当代经济学中的地位

经济学被定义为研究稀缺资源最优利用的科学, 在当代被分为微观经济学与宏观经济学两大部分. 近代主流微观经济学是所谓 “新古典学派”, 而主流宏观经济学是以凯恩斯为代表的有效需求理论还糅合进一点弗里德曼的货币学说. 不过至今微观经济学与宏观经济学尚不能整合成一个系统的理论. 这里我们的讨论局限

于微观经济学.

20 世纪二、三十年代以萨缪尔逊为代表的一批经济学家承继 19 世纪李嘉图 - 穆勒 - 马歇尔 - 瓦尔拉斯的传统, 大量引进微积分、数理统计、线性代数等数学工具构建新的经济学, 他们的目标是证明市场的自我调节会达到均衡, 这是一种最有效的状态, 能实现福利社会. 总之他们的目的是论证市场经济制度是最好的经济制度.

最后, 在五、六十年代, 以阿罗和德布鲁的工作为代表, 他们终于在公理化的基础上完成这个架构.

在这个理论中, 市场经济即商品经济由三部分构成: 消费者, 生产者与市场, 因此其理论由消费者行为, 生产者 (厂商) 行为, 竞争市场的一般均衡理论 (价格体制下) 构成, 再加上福利经济学作为辅助.

消费者行为被描述为在预算约束下根据个人偏好作出决策, 选购商品, 达到效用最大化.

因此在新古典经济学的消费者行为理论中, "偏好" 是最基本的出发点.

效用函数则是偏好的数量化. 数理经济学为证明可用效用函数最优化以作商品选择, 有非常严格而细致的讨论.

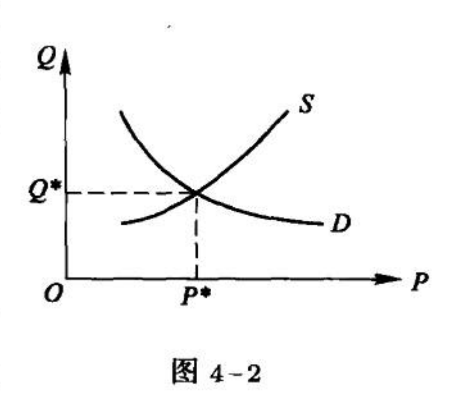


图 4-2

这样一来就可以建立起需求函数 D, 配上从厂商利润最大化中导出的供给函数 S, 就能证明一般均衡的存在性 (既决定价格 P^*, 也决定了数量 Q^*).

这些理论构成数理经济学这一新学科, 其数学工具也升级为凸分析和拓扑学.

五、期望效用在不确定性经济学中的作用

传统经济学假定消费者与生产者的决策所产生的结果是肯定而唯一的, 这显然不符合实际, 这就促使了不确定性下的经济学的发展. 这种要求在投资过程和金融行业 (银行, 证券, 保险) 中最为迫切. 事实上, 不确定经济学的应用主角也是金融学, 形成了金融经济学. 任何投资都是以今日的确定性投入以博取明日更大的但不确定性的回报, 这里就有风险, 并且要求定量化——处理不确定性, 目前概率论还是唯一合适的数学工具.

金融学中的经济人称为投资者 (买股票的人与卖股票的人都是投资者, 证券公司、保险公司、银行只是中介机构), 因此投资者的行为是最需要研究的, 而投资者的行为主要取决于他们对风险的态度. 到目前为止, 期望效用函数是解决这个问题的最佳良方. 这样一来你就可以读懂丹尼尔・伯努利文章的题目了. 我们也就不能不佩服丹尼尔在 200 多年前所表现出来的远见卓识了. 丹尼尔・伯努利和克拉默的主要兴趣在于阐明期望效用优于期望货币值. 特别地, 他们证明了基于风险规避假设 (等价于递减的货币边际效用), 期望效用方法既可用来解释人们购买保单, 又可用于解决圣彼得堡悖论.

下面我们将介绍现代效用理论.

六、期望效用函数与投资者行为

投资问题面对不确定因素, 最终收益是随机变量, 它是 ω 的函数. 如果有两种投资途径, 其最终收益分别是 $X(\omega)$ 与 $Y(\omega)$, 对某些 ω, $X(\omega) > Y(\omega)$, 对另一些 ω, 则可能相反. 怎么比较, 如何选择?

当代经济学家的解答是找一个效用函数 $U(x)$, 再利用期望效用函数 $EU(x)$ 来作决策: 若 $EU(X) > EU(Y)$, 则选择有更高期望效用的 X.

对于给定的效用函数 U, 通过 $U=C$ 定义**无差异曲线**, 这样定义的无差异曲线 ① 两两不相交; ② 两无差异曲线之间还有另一无差异曲线; ③ 若 U 是 x, y (两种商品消费量) 的二元函数, 则在坐标系中, 右上的无差异曲线比左下的好. 无差异曲线描述偏好, 它们与预算约束线的切点确定了最佳 (商品) 选择.

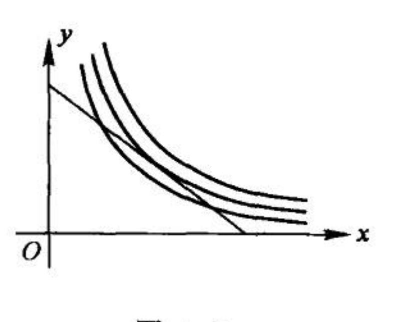


图 4-3

1944 年 von Neumann & Morgenstein 出版《博弈论与经济行为》, 这是 20 世纪影响数学与经济结合最重要的一本书, 承继 200 年前丹尼尔·伯努利的创意在公理化的基础上建立了期望效用理论, 证明了如果决策者的行为满足一组合理的一致性条件, 期望效用可得出非确定性条件下的最佳结果, 更详细地说, 他们证明了效用可引入到决策问题上, 以致那些基于期望效用作出决策的人等价于依照其偏好作出决策.

七、效用函数与对风险的态度

若投资者的效用函数为 $U(x)$, 则 $X \sim \begin{pmatrix} x_1 & x_2 \\ p_1 & p_2 \end{pmatrix}$ 的期望效用为

$$p_1U(x_1)+p_2U(x_2)$$

而 X 的期望值 $EX=x_1p_1+x_2p_2=\overline{x}$, 其效用为

$$U(p_1x_1+p_2x_2)$$

若

$$U(p_1x_1+p_2x_2)>p_1U(x_1)+p_2U(x_2) \tag{7}$$

则称投资者为**风险规避者**.

(7) 等价于要求效用函数 $U(x)$ 为上凸函数, 即满足 $U'(x)>0$, $U''(x)<0$.

若

$$U(p_1x_1 + p_2x_2) < p_1U(x_1) + p_2U(x_2) \tag{8}$$

则称投资者为**风险爱好者**.

(8) 等价于要求效用函数 $U(x)$ 为下凸函数, 即满足 $U'(x) > 0$, $U''(x) > 0$.

若

$$U(p_1x_1 + p_2x_2) = p_1U(x_1) + p_2U(x_2) \tag{9}$$

则称投资者为**风险中立者**.

(9) 等价于 $U(x)$ 为线性函数.

上述讨论中起关键作用的是延森不等式, 可参看《概率论基础》引理 4.3.2. 讨论中虽对二值变量进行, 实可推广至取有限个值的普遍情况.

值得说明, 效用函数不唯一. 若 $U(x)$ 为效用函数, 则 $aU(x)+b$, $a > 0$ 可表示同一偏好.

经验调查研究表明, 绝大部分人是风险规避者: 观其效用函数, 首先, $U'(x) > 0$, 这表明财富越多越好, 人人如此; 其次 $U''(x) < 0$, 即表示边际财富效用递减——穷人多得 100 元比富人多得 100 元有用得多.

边际效用递减是经济学中的一条基本原理.

凡满足 $U'(x) > 0$, $U''(x) < 0$ 者为风险规避者, 大多数人都是风险规避者.

值得注意丹尼尔·伯努利所用的对数效用函数和克拉默所用的平方根效用函数都是上凸函数, 而且他们对期望效用理论的解释与现代几乎一模一样, 怎能讲 "它没有科学的意义" 呢?

*八、确定性等价与风险溢价

对于上凸效用函数, 即对于风险规避的投资者, 风险收益 X 的效用不如平均收益 $\overline{x}$ 的效用, 见图 4-4. 这时可找到一个数 x^*, 它

满足 $x^* < \overline{x}$, 但

$$U(x^*) = p_1 U(x_1) + p_2 U(x_2) = EU(X) \tag{10}$$

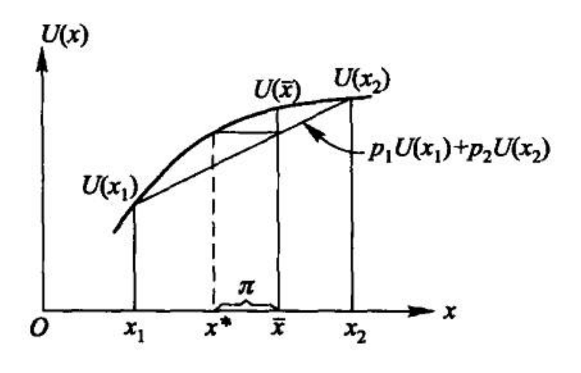


图 4-4

称 x^* 为随机收益 X 的**确定性等价**. 称 $\pi = \overline{x} - x^*$ 为**风险溢价**, 意思是只有额外增补 π 才愿参加该项有风险的随机游戏, 即为达到公平博弈必须的补偿.

举个例子, 假如一笔款存银行到期可得 1 万元, 而投资一个项目, 期望所得也是 1 万元, 则无人要投这个项目, 除非另有补偿.

一般地, 若对一切财富 W 成立

$$EU(W + X) = U(W + EX - \pi) \tag{11}$$

则称 π 为 U 的**风险溢价**.

Pratt 证明了

$$\pi \approx -\frac{D(X)}{2} \cdot \frac{U''(W)}{U'(W)} \tag{12}$$

阿罗证明了, 对 $X \sim \begin{pmatrix} -h & h \\ 1-p & p \end{pmatrix}$,

$$p - \frac{1}{2} \approx -\frac{h}{4} \cdot \frac{U''(W)}{U'(W)} \tag{13}$$

因此 $-\dfrac{U''(W)}{U'(W)}$ 与风险溢价成正比, 可作为风险的测度.

称 $A(W)=-\dfrac{U''(W)}{U'(W)}$ 为**绝对风险规避测度**.

称 $R(W)=-\dfrac{U''(W)}{U'(W)}W$ 为**相对风险规避测度**.

本教学札记介绍的内容为读者进入金融经济学所必备.

教学札记之二十一

现代投资理论和概率论

一、前言

现代投资理论是在马科维兹 (H. Markowitz) 1952 年提出的**证券组合选择理论**的框架上发展起来的. 1958 年托宾 (J. Tobin) 引入无风险资产并发现二基金分离定理. 1960 年代, 威廉・夏普 (William Sharpe), 约翰・林特纳 (John Lintner) 和简・莫辛 (Jan Mossin) 分别独立地建立了**资本资产定价模型** (CAPM), 该模型用于描述证券的风险与期望收益率的关系, 是测量风险、估价证券价值的基准和衡量投资绩效的标准. 随后罗斯 (Stephen Ross) 于 1976 年提出了**套利定价理论** (APT). 这些模型用均值和方差分别描述投资中的收益和风险, 并建立它们的**二阶矩 (即均值 - 方差) 模型**, 使投资学在概率统计的基础上建立了自己的理论框架, 并很快得到广泛而一致的承认, 如今已成为行业的主要语汇, 并指导着实际资本市场.

二、证券组合选择

1952 年马科维兹 "组合选择" 一文在 Journal of Finance 上发表, 这篇 14 页的论文是他的博士论文的摘要. 马科维兹是芝加哥大学经济系的博士生, 在校期间他参加了考尔斯经济委员会的工作. 这个委员会由考尔斯 (Cowles) 家族资助, 其创始人 A・考尔斯是投资领域定量分析的开拓者之一, 并创办了 Econometrica 杂志, 几乎全美的诺贝尔经济学奖得主都曾在这个委员会做过研究, 对经济

计量学和统计学在美国的发展做出过不可磨灭的贡献. 马科维兹受到委员会两任主持人 Koopmans 和 Marschak 的指点, 开始以股票市场作为研究主题, 并被介绍给当时的院长, 金融学家 Ketchun, 后者指导他读当时最好的一本投资学著作, 威廉斯的《投资价值理论》, 该书的要点是所谓"红利贴现模型". 有一天马科维兹脑中"突然浮现一个观念, 人们可能不只是重视报酬, 同时也在意风险". 这样马科维兹抓住问题的要害: 风险是整个投资过程的核心.

当时费勒的名著《概率论及其应用》第一卷刚问世, 许多美国经济学家都是通过这本书熟悉概率论语言, 并与此书终生为友.

马科维兹把收益率 r 看作随机变量, 假定投资者将追求期望收益率 $\bar{r}$ 最大化和方差 σ^2 最小化的二元目标, 把投资化为一个选择问题. 从数学上看是一个在线性约束下的二次规划问题, 这在《概率论基础》第四章 §2 例 13 中已有介绍.

马科维兹的论文在博士论文答辩中遇到麻烦, 著名货币学派创始人弗里德曼提出: 怎么能对这样一篇不是经济学, 也不是数学, 甚至不是企业管理论文的作者授予经济学博士学位呢? ——当然论文还是通过了.

但是这篇论文发表后并未受到重视, 直至另一位大经济学家托宾关注并邀请马科维兹到那时已搬迁到耶鲁大学的考尔斯委员会工作一年, 在那里他写成一本书《资产选择——投资的有效分散化》, 1987 年又写成系统的著作《资产组合选择和资本市场的均值-方差分析》. 但是总的说, 马科维兹已把他的研究兴趣转向运筹学和计算机软件, 因此当 1990 年被授予诺贝尔经济学奖时他本人也有点意外.

50 多年过去, 马科维兹的理论已被演绎成很漂亮的理论, 下面只能做最简单的介绍.

三、最小方差组合与有效前沿

基本假定: 单期、无摩擦、竞争市场. 投资者对财富不厌足, $\bar{r}_p$

越大越好; 投资者规避风险, σ_p^2 越小越好. 在《概率论基础》第四章§2 例 13 中, 投资组合问题归结为如下二次规划问题:

$$\min \quad \sigma_p^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij}$$

$$\text{s. t.} \quad \sum_{i=1}^{n} x_i \overline{r}_i = \overline{r}_p \tag{1}$$

用拉格朗日乘数法可找到这个问题的解, 单个称**最小方差组合**, 全体称**最小方差组合前沿**, 其上半段称为**有效前沿**——最好组合只需在有效前沿中寻找 (图 4-5).

在 $(\sigma_p, \overline{r}_p)$ 平面上的图形为

$$\frac{\sigma_p^2}{\frac{1}{C}} - \frac{\left(\overline{r}_p - \frac{A}{C}\right)^2}{\frac{BC - A^2}{C^2}} = 1 \tag{2}$$

其中 A, B, C 是由 n 种证券的收益率 $r_1, \cdots, r_n$ 的均值向量 $(\overline{r}_1, \cdots, \overline{r}_n)$ 和协方差阵 (σ_{ij}) 所决定的常数, 它们是这个问题的已知量. 这是双曲线的一支.

有了表达式就可对最小方差组合作很细致的讨论和得出许多重要结论, 此处从略.

四、无风险证券与二基金分离定理

托宾提出应把无风险资产 (国债, 银行存款等) 纳入投资组合这十分合理, 而且把权益性股票与固定收益的债券一道研究, 建立统一的证券投资理论也十分有益.

无风险证券 (利率为 r_f) 的引入使整个理论更加完满. 这时新的有效前沿是从 r_f 出发的一根射线, 它与风险资产的有效前沿相切于 T 点 (图 4-6).

这样一来, 投资者为得到最好投资组合只需把资金在二个证券组合 (一个是无风险证券, 一个是切点证券组合) 之间进行分配.

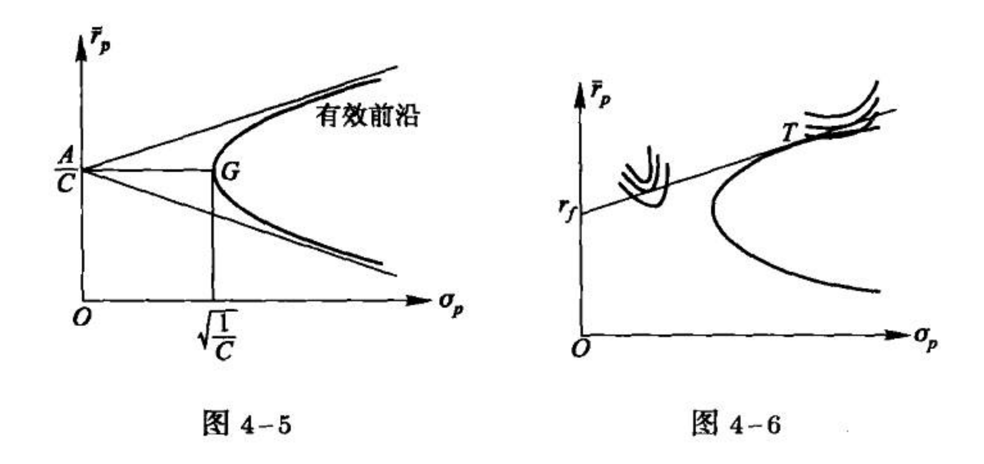


图 4-5　　　　图 4-6

至于如何分配, 则与每个投资者对风险的偏好程度 (用上凸效用函数表示) 有关, 利用无差异曲线与有效前沿相切来寻找最好组合. 对风险很厌恶的人取 T 点左边的组合, 这时无风险证券比例大; 风险爱好者则甚至贷款来炒股, 选择 T 点右边的组合.

因此在有无风险证券存在的场合, 无论投资者最终投资于何种有效组合 p, 它在风险资产中各种证券所占的比例固定不变, 恒为切点组合 T 的投资组合比例, 因此切点组合 T 成为**市场组合** M. 这就是二基金分离定理的最早的形式, 也是最简单的形式, 它可以推广到一般场合.

事实上, 在有无风险证券存在场合, 分配给 n 种证券的资金份额 $x_1, \cdots, x_n$, 其和小于 1, 而 $1 - x_1 - \cdots - x_n$ 则用于无风险证券, 这时最优化问题表述如下:

$$\begin{aligned} \min \quad & \sigma_p^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij} \\ \text{s. t.} \quad & \sum_{i=1}^{n} x_i \bar{r}_i + (1 - x_1 - \cdots - x_n) r_f = \bar{r}_p \end{aligned} \tag{3}$$

用拉格朗日乘数法可得

$$\sum_{j=1}^{n} \lambda x_j \sigma_{ij} = \bar{r}_i - r_f, \quad i = 1, 2, \cdots, n \tag{4}$$

这里

$$\lambda=\frac{\overline{r}_T-r_f}{\sigma_T^2}$$

其中 $\overline{r}_T$ 及 σ_T^2 分别为切点组合 T 的均值与方差.

由上式可解出 $x_1,x_2,\cdots,x_n$ 以确定最好组合.

以上对马科维兹组合选择理论作了近代表述. 从中可以看出, 这个理论兼顾了投资中所关心的收益与风险两大要素, 对证券组合的分散化功能作出有说服力的说明. 在一些基本的假定下可找出最好投资方案, 理论和应用上都有重大价值.

进一步讲, 马科维兹事实上建立了不确定性经济中投资者的行为理论, 因此在他获诺贝尔经济学奖的演讲中说: "至于他 (弗里德曼) 的论点的是非, 此时我很愿意让步: 在我答辩论文时证券组合理论不是经济学的一部分, 但是现在它是. "

顺带指出, 马科维兹的双目标决策模型并不能完全纳入新古典经济学的理论框架, 因此甚至马科维兹在上述演讲中也花不少篇幅来说明, 如果适当选择效用函数, 例如 $U(r)=\log(1+r)$, 它与用下式表示的目标函数 $f(\overline{r},\ \sigma^2)=\log(1+\overline{r})-5\sigma^2/(1+\overline{r})^2$ 能得到近似相同的结论. 但是, 事实上, 马科维兹的模型在实际运作中更切实可用.

另一个问题是以方差作为风险度量, 连首创者都存一定疑虑, 但是半个世纪过去, 方差作为证券风险的度量的地位, 不但不动摇, 甚至更稳固. 试想在每天海量的证券交易记录面前, 不用方差还有什么量更能度量 "风险". 从另一角度看, 如果有谁能发现一个比方差更好的度量随机变量离散程度的指标, 岂不带来概率论的新的一轮大发展!

五、夏普的资本资产定价模型

夏普在 20 世纪 50 年代是 UCLA 的学生, 主修经济学也选修金融学. 当他写完博士论文提纲交给新指导老师 Hirshleigh 时, 这

位当时赫赫有名的教授一周后把论文退回并表示没有什么内容可言. 夏普把这件事告诉他的另一位老师 Waston, 后者建议他找马科维兹谈谈. 长久以来 Waston 一直是马科维兹著作的热情拥护者和传播者.

当时马科维兹在 RAND 公司工作, 该公司位于洛杉矶的圣莫尼卡, 离 UCLA 不远. 夏普找到马科维兹之后, 发现他们的兴趣有交集. 随后校方同意夏普的申请让他在马科维兹的指导下完成研究生学业.

马科维兹深知自己理论的弱点, 首先是计算量太大, 对数目以百、千计的股票种类 n, 要搜集它们的期望收益 $\bar{r}_i$, 特别是协方差阵 (σ_{ij}), 在 20世纪 50 年代计算条件下根本无法完成, 要及时解出方程 (1), (3) 也很困难. 因此他一直在研究简化的模型, 考虑到最好组合都只与市场组合 M 有关, 因此他提出了**单因子模型**. 这时夏普找到了他, 从此之后, 他就退居指导者的地位, 由夏普把理论向前推进. 夏普也的确作出重大贡献, 不断有新的创意出来. 更重要的是他一直在金融学领域执教和研究, 写了几本很有影响的专著和教科书, 对推广普及这套理论很有作用, 因此当他在 1990 年与马科维兹同获诺贝尔经济学奖时, 他的名气大得多, 早就做过美国金融学会主席.

下面来推导 CAPM (Capital Asset Pricing Model) 模型.

改写 (4) 为

$$\lambda \operatorname{cov}\Big(r_i, \sum_{j=1}^{n} x_j r_j\Big) = \bar{r}_i - r_f, \quad i = 1, 2, \cdots, n \tag{5}$$

或

$$\bar{r}_i = r_f + \lambda \operatorname{cov}\Big(r_i, \sum_{j=1}^{n} x_j r_j\Big), \quad i = 1, 2, \cdots, n \tag{6}$$

所有最好组合都有相同的权 $x_1, x_2, \cdots, x_n$, 因此若以 $\bar{r}_M$ 与 σ_M^2 记

市场组合 M 的均值和方差, 则

$$Er_M = \overline{r}_M = \sum_{j=1}^{n} x_j \overline{r}_j \tag{7}$$

又

$$\lambda = \frac{\overline{r}_M - r_f}{\sigma_M^2} \tag{8}$$

故

$$\begin{aligned} \overline{r}_i &= r_f + \frac{\overline{r}_M - r_f}{\sigma_M^2} \cdot \mathrm{cov}(r_i,\ r_M) \\ &= r_f + \beta_{iM}(\overline{r}_M - r_f) \end{aligned} \tag{9}$$

其中

$$\beta_{iM} = \frac{\mathrm{cov}(r_i,\ r_M)}{\sigma_M^2} \tag{10}$$

(9) 是对第 i 个证券 S_i 的收益率 r_i 的 CAPM, 利用对 r_i 的线性性, 可知对任何证券组合 p 的收益 r_p, 均有

$$\overline{r}_p = r_f + \beta_{pM}(\overline{r}_M - r_f) \tag{11}$$

其中

$$\beta_{pM} = \frac{\mathrm{cov}(r_p,\ r_M)}{\sigma_M^2} \tag{12}$$

这是对一般证券组合 p 的 CAPM.

(9) 表明, 为让投资者购买第 i 种证券 S_i, 它的报酬即收益应由两部分组成: 其一是 r_f, 相等于存银行得到的利息; 另一部分是 $\beta_{iM}(\overline{r}_M - r_f)$. 其中 $\overline{r}_M - r_f$ 表示市场组合的超额收益, 它乘以 β_{iM} 构成的第二部分 $\beta_{iM}(\overline{r}_M - r_f)$ 则是购买风险组合 M 因冒风险得到的报酬, β_{iM} 是证券 S_i 与 M 的相关指标, 称为 **β 系数**.

资本资产定价模型是一般均衡性理论, 通过市场均衡给各种证券定价. 业绩很差的股票价格会下跌, 从而使它的收益率达到 CAPM 要求的均衡值, 业绩好的股票收益高, 价格会上涨, 这时购买它的成本上升, 收益率会降到均衡水平.

资本资产定价模型主要由三个线性方程构成:

(i)
$$\overline{r}_p = r_f + \frac{r_M - r_f}{\sigma_M}\sigma_p \tag{13}$$

称为**资本市场线** (Capital Market Line), 它表明市场组合 M 也是有效组合, 所有有效组合均在射线上.

(ii)
$$\overline{r}_p = r_f + \beta_{pM}(\overline{r}_M - r_f) \tag{14}$$

即 (11) 式, 称为**证券市场线** (Security Market Line). 如果单讲 CAPM 的话就是指这条线, 这时衡量风险的指标是 β.

(iii)
$$r_i - r_f = \alpha_i + \beta_{iM}(r_M - r_f) + \varepsilon_i, \quad i = 1, 2, \cdots, n \tag{15}$$

其中设 $\varepsilon_1, \cdots, \varepsilon_n$ 互不相关, 均服从 $N(0, \sigma^2)$, $\mathrm{cov}(r_M, \varepsilon_i) = 0$.

(15) 称为**特征线**, 对每个公司可用实际资料估计 α_i, β_{iM}.

对 (15) 两边取方差得

$$\sigma_i^2 = \beta_{iM}^2\sigma_M^2 + \sigma_{\varepsilon_i}^2 \tag{16}$$

它把每个证券的风险分成两部分: 第一部分 $\sigma_{\varepsilon_i}^2$ 称为**个别风险**, 第二部分 $\beta_{iM}^2\sigma_M^2$ 称为**市场风险**或**系统风险**. 个别风险可通过组合分散化消除, 而市场风险只对 β_{iM} 带来的风险提供报酬, 而 $\dfrac{\overline{r}_M - r_f}{\sigma_M}$ 称为**风险的市场价格**.

CAPM 是很实用的理论. 这里不予详述.

六、罗斯的套利定价理论

1976 年罗斯提出套利定价理论, 这个理论有两大要素, 其一是多因子模型, 其二是套利定价概念.

前已提及, 是马科维兹最早研究单因子模型, 后来夏普继续他的研究, 他所称的 "对角线模型" 实际上也是单一因子模型. 用单一因素 (市场组合 M) 来解释各个股票价格的变动过于粗略而且实用效果也不太好, 因此罗斯提出多因子模型使理论有更大适用范围. 顺便说, 美国著名金融学家 Fama 数十年研究的结果最后推荐的是一个三因子模型.

套利定价的提出意义重大. 罗斯在论文最后对套利定价理论的解释也引起许多讨论. 不过我们将把套利定价理论放在教学札记之二十二 "数理金融学和概率论" 中介绍.

罗斯讨论的模型是

$$r_i = a_i + b_{i1}F_1 + b_{i2}F_2 + \cdots + b_{ik}F_k + \varepsilon_i, \quad i = 1, 2, \cdots, n \qquad (17)$$

假定诸 ε_i 互不相关, 均服从 $N(0, \sigma^2)$; 诸因素 $F_1, F_2, \cdots, F_k$ 也互不相关; F_j 与 ε_i 也互不相关.

在无套利假定下, 结论是

$$\overline{r}_i = r_f + \sum_{j=1}^{k} b_{ij}\lambda_j, \quad i = 1, 2, \cdots, n \qquad (18)$$

其中 λ_j 是风险因素 F_j 的市场价格.

理论断言, 承担一份风险就有一份超额收益, 得到与 CAPM 一致的结论. 显然罗斯的多因子模型把单因子模型作为特例, 但他研究的无套利定价与 CAPM 研究的均衡定价不同, 因此两个模型互不包含.

值得指出, CAPM 与 APT 均与线性回归关系密切.

教学札记之二十二

数理金融学和概率论

一、从布朗运动到随机分析

1900 年巴舍利耶 (L. Bachelier, 1870—1946) 在庞加莱指导下完成他的学院派博士论文《投机理论》, 他率先使用概率理论, 包括各种数学工具来解释股票市场的运作, 而且占了整本论文约 70 页的篇幅. 巴舍利耶的思想远远跑在时代前面. 在金融方面, 他不但研究股票还研究期货与期权等金融工具, 提出它们的定价理论, 还分析巴黎股票市场 1894—1898 年的数据. 在数学方面, 他第一个提出了连续时间随机过程的概念, 并研究了布朗运动. 须知从物理学角度给布朗运动以正确定性解释和定量描述的是爱因斯坦于

1905 年发表的著名论文和 1906 年波兰物理学家斯摩罗霍夫斯基的论文. 美国最有名的经济学家萨缪尔逊曾说, 他详细读过巴舍利耶与爱因斯坦的论文并作比较, 发现从各个角度看巴舍利耶的论文都更胜一筹.

在巴舍利耶时代要严格描述随机过程, 超前太多, 虽然他的工作给出了马尔可夫性和切普曼－科尔莫戈罗夫方程, 用 4 种方法证明分布的正态性, 还证明方差为 $\sigma^2 t$ 等非常深刻的结果, 甚至用了反射法, 但是数学上不严格或不正确之处也存在. 在经济方面, 他以正态分布描述股票价格, 其取负值的可能性违背了 "有限责任" 的原则, 在期权定价中带来荒谬的结论. 这一切的结果使他难以获得好的工作职位, 该文在经济界被埋没几达 60 年之久.

幸而数学界并没有忘记巴舍利耶的工作, 从 1918 年起美国数学家维纳 (N. Wiener, 1894—1964) 建立了布朗运动的数学理论, 因此布朗运动在概率论文献中有一个同样通行的名字——维纳过程. 维纳过程和泊松过程是最重要的两种具体随机过程, 是每一个研究随机过程的人所熟悉的.

对布朗运动的进一步研究在法国本土由莱维继续. 莱维是 20 世纪 20 ~ 30 年代最有创意的概率学家, 除了在古典中心极限定理和特征函数连续性定理中有重大贡献外, 是他第一个把独立和与独立增量过程的研究结合起来, 导致无穷可分分布律研究的完成, 对布朗运动的深入研究和鞅序列概念的提出也是他的重大成就. 莱维思想十分丰富, 但他的表达很奇特, 他的书常人很难读懂, 导致后继无人. 日本的伊藤清 (K. Itô, 1915—2008) 受到莱维启发在 1942 年发明随机积分, 后又发展随机微分方程, 成为 Itô 微积分, 是现代随机分析中最基础也最有用的部分. 伊藤清因此获 1987 年沃尔夫数学奖, 2006 年更获第一届高斯奖.

鞅序列由伯恩斯坦首先系统研究, 由 Ville 定名, 最后由美国杜布加以发展成为系统的理论. 战后法国梅耶到美国学回概率论, 以他与杜布证明的杜布－梅耶分解为起点发展成现代鞅轮.

随机积分 1967 年由伊藤清的两位学生国田和渡边推广到平方可积鞅, 并发展为现代随机分析. 鞅和随机分析形成概率论重要分支.

因此当经济学界重新发现巴舍利耶的工作, 并用来建立金融学的近代理论时, 发觉数学家早就为他们准备了充分的数学工具.

现代普遍的看法是巴舍利耶 1900 年的论文同时创立了两个新学科——连续时间随机过程和连续时间金融学, 而后面二者实际上是双胞胎.

二、关于股票市场价格随机特征的大辩论

1954 年美国统计学家 Savage 发现了巴舍利耶 1914 年的小册子, 立刻寄发明信片给经济学家询问他们是否知道这个人及其工作. 当时萨缪尔逊正着手探讨市场行为理论并构建自己的定价模型, 在 MIT 图书馆找不到这本小册子却发现巴舍利耶 1900 年论文副本, 从此经济学界知道了巴舍利耶的工作.

概率方法进入金融学最基本的问题是: 股价是否可测?

1933 年 A. 考尔斯在大量市场实际数据分析的基础上发表论文 "股市预测指标能够准确预测吗? " 结论是 "这是值得怀疑的."

1934 年 Warking 发表一篇文章认为 "价格变动倾向高度随机性". 他发现股价走势图与随机数产生的图形无法分辨.

1953 年英国著名统计学家 M. G. Kendall 对 1928—1938 年的大量英国市场数据, 1883—1934 年芝加哥小麦月平均价格, 1816—1951 年纽约棉花资料作分析得出与 Warking 同样的结论.

1965 年萨缪尔逊发表论文 "证明预期价格随机涨落", 同年法马 (E. Fama) 用计算机分析市场数据发表论文: "股票市场价格的行为", 都论证股价的随机性. 之后出了一本厚厚的 Cootner 主编的论文选集 "The Random Character of Stock Market Prices" revised ed. 1967.

最后由法马提出 "市场有效假设" (EMH) (相当于 "股票价格

的随机游动理论") 展开了至今长达 40 年的大论战, 发表论文不下千篇. 虽然很难得到完全一致的认识, 但股票价格之难于预测恐怕是谁都否定不了的事实.

正是这场大论战为当代数理金融学的建立清扫地基, 复归巴舍利耶的论题.

三、从几何布朗运动到期权定价公式

为克服巴舍利耶理论在经济方面的困难, 关键一步是用几何布朗运动代替巴舍利耶理论中的 (算术) 布朗运动. 在离散场合, 相当于假定股价变化率 ΔP 服从随机游动, 这与 Warking 和 Kendall 的结论完全一致, 这样当用随机游动逼近布朗运动 (见教学札记之九 "随机游动三题"), 股票的连续收益率服从正态分布而股价则服从对数正态分布.

采用了这个十分合理的假定之后, 一切经济学的困难都迎刃而解: 股价不再可能为负, 期权定价不会出现无穷.

接下来的开拓性工作由萨缪尔逊及其杰出的学生默顿 (R. Merton) 完成.

Merton 原在加州理工学院学习应用数学, 突然想改学经济学, 便向各校发信, 只有 MIT 要他. 入学后系主任建议他听萨缪尔逊的课, 从此他便成为萨缪尔逊的学生, 后来成为助手, 再后来成为合作者, 最后他超过他的导师成为数理金融学的主要开拓者. 关键之一是他掌握随机微积分.

当时主要有两个问题: 一个是最优消费与投资组合问题, 考虑如何在预算约束下在一个时期内安排消费与投资组合, 使效用达到最大. 这个问题比马科维兹模型在两个方向有了扩充, 一是既考虑投资又考虑消费; 二是从单期模型到多期模型. 最初是萨缪尔逊自己做 (离散的) 多期模型, 后来便与 Merton 合作做更深入的结果, 最后是 Merton 把它推广到连续时间模型, 充分利用随机微积分, 结果比离散场合简洁得多, 成了 Merton 的重大贡献.

这个方向后由黄奇辅, Duffie 推广到多维扩散过程描述的证券市场, 做出了许多重要理论成果. 这部分由于太理论, 太难, 还没有找到有效的应用, 后继研究者不多.

另一个问题是期权定价. C. W. Smith 1976 年曾写过一篇文章对这个问题的前期情况作过评述. 总的说, 不少人 (包括萨缪尔逊) 接近答案但都有重大问题, 直至 1972 年 MIT 的两个青年教师 F. Black 及 M. Scholes 才解决这个问题, 找到正确的公式, 但是他们的文章发表不出来, 知道他们结果而立即写出一篇更深研究结果的同事 Merton 只得也搁下来未发表, 反而是一篇由前二人撰写的实证性文章在另一杂志提前发表, 最后才是 1973 年两篇主要文章的正式发表. 这个公式现称 Black - Merton - Scholes 公式, 被认为是社会科学中应用数学获得的最高成就, 不但理论上深刻, 而且在指导着期权市场的每日实践. Merton 与 Scholes 因这个工作获 1997 年诺贝尔经济学奖, Black 不幸于前一年病逝.

下面我们通过这个公式来介绍金融经济学的主要经济概念和所用的数学方法.

经济概念包括: 无套利假设, 一价定律, 风险中立定价.

数学方法主要是 Itô 微积分.

四、套利定价

当代金融产品的定价分为两大类: 一类是无条件定价, 亦称因素定价, CAPM 及 APT 属于此类; 另一类是条件定价, 用的是无套利定价方法, 期权与其他衍生金融产品的定价属于此类. 本小节介绍后一类定价的经济学原理.

1. 一价定律

在一个竞争的市场中, 如果两个资产是等价的, 它们将有相同的市场价格.

一价定律通过一个称为 “套利” 的过程来实现, 即利用差价, 以低价买进并立即以高价卖出等价产品, 以锁定一个确定性的利

润而不冒任何风险. 这时低价的必涨价, 高价的便跌价达到一价为止. 这是 MM (Modigliani - Miller) 定理的要点.

一价定律是金融学中最基本的定价原则.

推论 1: 利率是单一的.

推论 2: 股票在不同地点是同一价格的.

2. 套利机会

套利机会 A: 一项资产在时刻 0 得到正的收入, 但不带来将来的盈亏, 则称有 A 型套利机会.

市场不应有套利机会 A, 因此有如下推论:

推论 1: 价格是线性的.

推论 2: 若 $X_T = Y_T$, 则 $p(X_t) = p(Y_t)$, $0 < t < T$, 即一价定律.

套利机会 B: 一项资产, 在 0 时刻费用为 0, 但将来所得 X, 满足 $P\{X \geqslant 0\} = 1$, $P\{X > 0\} > 0$, 则称有 B 型套利机会.

满足上述两式的资产 X, 称为**有限责任资产**. 股票, 债券, 彩票等均是有限责任资产.

推论 3: 有限责任资产其价格为正.

占优定理: 若 $X \geqslant Y$, 则 $p(X) \geqslant p(Y)$; 若 $X > Y$, 则 $p(X) > p(Y)$.

五、二叉树定价

以无红利股票的买入期权的定价问题为例.

股票现价 20, 三月期, 敲定价 $K = 21$, 无风险利率 $r = 0.12$, 以二叉树表示股价, 下面标明相应的期权价格. 见图 4 - 7.

这里是单期模型, 期末股价可升至 22, 期权价格为 $22 - 21 = 1$, 期末价格若降至 18, 期权作废, 价格为 0, 要求 0 时刻期权价格 f. 这时可构造如下组合: 买进 Δ 份股票, 卖出一份买入期权, 选 Δ 使这个组合无风险.

若股价上升, 组合的价值为 $22\Delta - 1$.

若股价下降, 组合的价值为 18Δ.

为使组合无风险, 必须有 $22\Delta - 1 = 18\Delta$, 求得 $\Delta = 0.25$, 这时不论股价升降, 组合的价值均为 4.5.

其现值为 $4.5 \times e^{-0.12\times\frac{3}{12}} = 4.367$.

从而组合的初始价值为 $20 \times 0.25 - f = 4.367$, 故 $f = 0.633$.

这个数值例子所用的方法可以一般化. 见图 4-8.

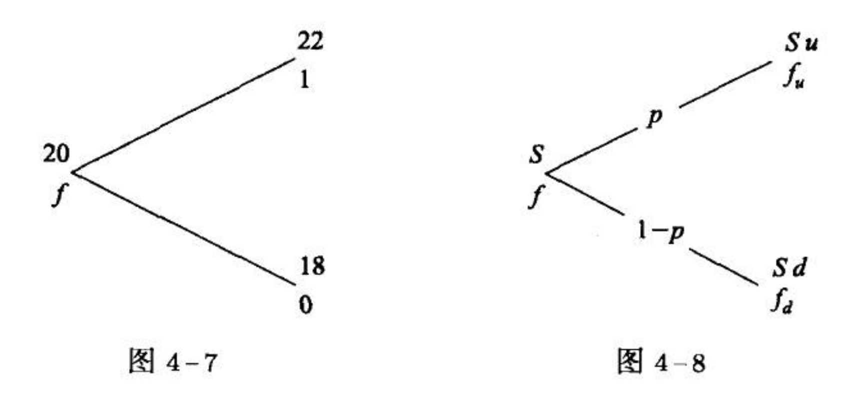


图 4-7　　　　图 4-8

若在时刻 0, 股票的价格为 S, 一种关于它的衍生证券的价格为 f.

在时刻 T, 股票价格上升到 Su 或下降到 Sd, 其中 $d < 1 < u$. 这时衍生证券的损益分别为 f_u 与 f_d. 因为股票和它的衍生证券 (例如买入期权) 有同一风险来源, 因此可以构成它们的一个证券组合, 互相抵消风险而成一个无风险资产. 按套利定价原则 (或一价定律) 这个组合的收益率应为无风险利率 r, 所建立的组合称为**套期保值组合**. 设有 Δ 份股票与卖出一份衍生证券, 在 T 时刻, 若股价上升, 该组合的价值为 $Su\Delta - f_u$; 若股价下降, 该组合的价值为 $Sd\Delta - f_d$. 为使组合无风险, 应有 $Su\Delta - f_u = Sd\Delta - f_d$, 从而

$$\Delta = \frac{f_u - f_d}{Su - Sd} = \frac{\text{衍生证券价格差}}{\text{股票的价格差}} = \frac{\Delta C}{\Delta S} \tag{1}$$

称为**套期保值比率**. 用此比率构成证券组合, 不论股价升降, 组合价值均为 $Su\Delta - f_u$ 或 $Sd\Delta - f_d$.

在 0 时刻, 建立这个组合的费用是 $S\Delta - f = (Su\Delta - f_u)\,e^{-rT}$,

因此衍生证券的价格为

$$f = S\Delta - (Su\Delta - f_u)\,\mathrm{e}^{-rT} = \mathrm{e}^{-rT}\big(qf_u + (1-q)f_d\big) \qquad (2)$$

其中 $q = \dfrac{\mathrm{e}^{rT} - d}{u - d}$.

在上例中, $u = 1.1$, $d = 0.9$, $r = 0.12$, $T = \dfrac{3}{12}$, $f_u = 1$, $f_d = 0$, $q = \dfrac{\mathrm{e}^{rT} - d}{u - d} = 0.6523$,

$$f = \mathrm{e}^{-0.12 \times \frac{3}{12}}[0.6523 \times 1 - 0.3477 \times 0] = 0.633$$

风险中立定价 Cox & Ross (1976) 指出在风险中立世界中, 一切投资者对风险无所谓, 从而有: (1) 一切资产都有同一收益率; (2) 价格可由未来期望值以利率 r 贴现算出. 二叉树中的人造概率 $q = \dfrac{\mathrm{e}^{rT} - d}{u - d}$ 即为**风险中立概率**.

这就提供了另一种定价思路.

以上做法可推广到多期模型. 这是 Cox – Ross – Rubinstein 在 1979 年提出的一个模型, 既能说理也能用于数值计算, 是期权定价中十分有名的模型, 实质上是伯努利试验的乘法模型, 其连续逼近为几何布朗运动.

六、Itô 微积分简介

若有 $f(t)$ 股证券, 其每股价格为布朗运动 $W(t)$, 则 $[0, T]$ 中的收益为 $\displaystyle\int_0^T f(t)\,\mathrm{d}W(t)$, 由于 $W(t)$ 处处连续处处不可微, 这种积分在通常意义下无法定义, Itô 用均方收敛来定义.

所谓随机微分方程:

$$\mathrm{d}S_t = a(S_t,\, t)\,\mathrm{d}t + \sigma(S_t,\, t)\,\mathrm{d}W_t \qquad (3)$$

意味着

$$\int_t^{t+\tau} \mathrm{d}S_t = \int_t^{t+\tau} a(S_t,\, t)\,\mathrm{d}t + \int_t^{t+\tau} \sigma(S_t,\, t)\,\mathrm{d}W_t$$

这种方程解的存在性与唯一性已证得，且为马尔可夫过程，通常称 **Itô 过程**.

Itô 微积分最基本的公式是

$$\mathrm{d}W_t \cdot \mathrm{d}W_t = \mathrm{d}t \tag{4}$$

Itô 公式: 若 $\mathrm{d}S_t = a_t\mathrm{d}t + \sigma_t\mathrm{d}W_t$, 设 $F(S_t, t)$ 关于 S_t 二次可微, 关于 t 一次可微, 则

$$\mathrm{d}F_t = \frac{\partial F}{\partial S_t}\,\mathrm{d}S_t + \frac{\partial F}{\partial t}\,\mathrm{d}t + \frac{1}{2}\,\frac{\partial^2 F}{\partial S_t^2}\,\sigma_t^2\,\mathrm{d}t \tag{5}$$

把 (3) 代入 (5) 得

$$\mathrm{d}F_t = \left[\frac{\partial F}{\partial t} + \frac{\partial F}{\partial S_t}\,a_t + \frac{1}{2}\,\frac{\partial^2 F}{\partial S_t^2}\,\sigma_t^2\right]\mathrm{d}t + \frac{\partial F}{\partial S_t}\,\sigma_t\,\mathrm{d}W_t \tag{6}$$

因此 F_t 的微分仍是 Itô 微分, F_t 仍是 Itô 过程.

七、Black－Scholes 公式的推导

基本假定:

无摩擦竞争市场, 股票价格服从几何布朗运动

$$\frac{\mathrm{d}S_t}{S_t} = \mu\mathrm{d}t + \sigma\mathrm{d}W_t$$

有无风险债券, 具常利率 r, 即 $\mathrm{d}B(t) = rB(t)\mathrm{d}t$;

证券在 $[0, T]$ 中连续交易, 市场无套利机会, 股票不派红利.

先用无套利方法, 构造一个无风险组合: Δ 份股票多头和一份衍生证券空头, 其价值为 $\pi = \Delta S - f$, 在 $[t, t+\Delta t]$ 中假定 Δ 不变, 用 Itô 公式,

$$\begin{aligned}
\mathrm{d}\pi &= \Delta\mathrm{d}S - \mathrm{d}f \\
&= -\left(\frac{\partial f}{\partial S}\,\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\,\frac{\partial^2 f}{\partial S^2}\,\sigma^2 S^2\right)\mathrm{d}t \\
&\quad - \frac{\partial f}{\partial S}\,\sigma S\,\mathrm{d}W_t + \Delta\mu S\mathrm{d}t + \Delta\sigma S\mathrm{d}W_t
\end{aligned}$$

为使 $\mathrm{d}W_t$ 的系数为 0, 必须而且只需取 $\Delta = \dfrac{\partial f}{\partial S}$——这就是**套期保值比率**, 这时组合无风险, 其利率必为 r, 故

$$\begin{aligned}\mathrm{d}\pi &= \Big(-\frac{\partial f}{\partial t} - \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\Big)\mathrm{d}t \\ &= r\pi\mathrm{d}t \\ &= r\Big(\frac{\partial f}{\partial S}S - f\Big)\mathrm{d}t\end{aligned}$$

即得 **Black – Scholes 方程**

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2 = rf \tag{7}$$

配合适当的边界条件, 解此偏微分方程, 可得 Black - Scholes 公式.

另一种解法是利用 Cox – Ross 的风险中立定价方法, 设股票的收益率为 r, 并以 r 作为贴现率计算价格, 这时计算公式为

$$C_t = \mathrm{e}^{-r(T-t)}\widehat{E}\big[\max(S_T - K, 0)\big] \tag{8}$$

这里期望 $\widehat{E}$ 对风险中立世界中的人造概率 Q 进行计算.

这正是我们在《概率论基础》习题四 **25 题中的做法, 具体计算过程见习题解答.

这种做法由 Harrison – Kreps (1979) 和 Harrison – Pliska (1981) 加以总结, 发现了与鞅有密切关系.

近代已证明:

资产定价基本定理 如下事实等价:

(1) 无套利; (2) 存在正的线性价格函数; (3) 存在等价鞅测度; (4) 存在随机正折现因子; (5) 最优证券选择问题有解.

九十年代的研究重点是利率期限结构, 本世纪初则为信用衍生品.

教学札记之二十三

事件的独立性与相关性

一、事件的独立性

事件 A 与事件 B 独立的定义为

$$P(AB) = P(A)P(B) \tag{1}$$

若事件 A, B 独立, 最直接的推论是

$$P(A|B) = P(A) \tag{2}$$

$$P(B|A) = P(B) \tag{3}$$

以后约定, 当写条件概率 $P(A|B)$ 时, 若未特别提出都假定 $P(B) > 0$.

事件 A, B 独立的另一推论是 $(\overline{A}, B), (A, \overline{B}), (\overline{A}, \overline{B})$ 也独立.

从 (2) 出发, 可以作进一步的讨论. 例如, 若成立

$$P(A|B) > P(A) \tag{4}$$

这时表明 B 的发生, 使 A 发生的条件概率增大, 因此可以说, B 的发生为 A 的发生带来 "正信息".

事实上, 若 $P(A|B) > P(A)$, 则

$$P(AB) = P(B)P(A|B) > P(A)P(B) \tag{5}$$

而且

$$P(B|A) = \frac{P(AB)}{P(A)} > \frac{P(A)P(B)}{P(A)} = P(B) \tag{6}$$

因此 A, B 有正信息是相互的.

同理, 若

$$P(A|B) < P(A) \tag{7}$$

必有

$$P(B|A) < P(B) \tag{8}$$

称事件 A 与事件 B 互相带来"负信息".

进一步, 若 $P(A|B) > P(A)$, 则

$$
\begin{aligned}
P(A|\overline{B}) &= \frac{P(A\overline{B})}{P(\overline{B})} = \frac{P(A) - P(AB)}{P(\overline{B})} \\
&< \frac{P(A) - P(A)P(B)}{P(\overline{B})} = P(A)
\end{aligned} \tag{9}
$$

因此若 B 给 A 带来"正信息", 则 $\overline{B}$ 必给 A 带来"负信息".

二、事件的相关性

《概率论基础》第四章 §2 引入事件 A 与事件 B 的相关系数

$$
\rho_{AB} = \frac{P(AB) - P(A)P(B)}{\sqrt{P(A)P(\overline{A})P(B)P(\overline{B})}} \tag{10}
$$

这是事件 A 的示性函数 $\mathbf{1}_A$ 与事件 B 的示性函数 $\mathbf{1}_B$ 的相关系数.

当 $P(A) = 0$ 或 $P(B) = 0$ 以及 $P(A) = 1$ 或 $P(B) = 1$ 时, 补充定义 $\rho_{AB} = 0$.

有了事件相关系数, 就可以讨论事件的相关性.

若定义事件 A 与事件 B 不相关为 $\rho_{AB} = 0$, 则

$$
\rho_{AB} = 0 \Longleftrightarrow P(AB) = P(A)P(B) \tag{11}
$$

因此事件 A 与事件 B 不相关的充要条件为事件 A 与事件 B 相互独立.

对于二值变量 ξ 与 η, 它们的不相关性与独立性等价. 这是与正态变量一齐分享的一个独特的性质, 必然具有很高的研究价值. 在正态场合, 主要好处是线性预测即为最佳预测. 关于这点在作者的另一本书中有较详细的讨论, 见《概率论与数理统计》. 复旦大学出版社, 2003.

这里我们关心的是把相关分析推广到事件场合.

当 $\rho_{AB} > 0$ 时称事件 A 与事件 B 正相关, 当 $\rho_{AB} < 0$ 时称事

件 A 与事件 B 负相关, 这时有如下关系:

$$\rho_{AB} > 0 \Longleftrightarrow P(AB) > P(A)P(B) \Longleftrightarrow P(A|B) > P(A) \quad (12)$$

$$\rho_{AB} = 0 \Longleftrightarrow P(AB) = P(A)P(B) \Longleftrightarrow P(A|B) = P(A) \quad (13)$$

$$\rho_{AB} < 0 \Longleftrightarrow P(AB) < P(A)P(B) \Longleftrightarrow P(A|B) < P(A) \quad (14)$$

(12) 对应于正相关即“正信息”的场合; (14) 对应于负相关即“负信息”的场合; (13) 自然是独立即不相关的场合, 这时 $\overline{A}$ 与 B, A 与 $\overline{B}$, $\overline{A}$ 与 $\overline{B}$ 也都不相关即独立.

特别地, $\rho_{AB} = 1$, 即完全正相关场合, 对应于 $A = B$; 而 $\rho_{AB} = -1$ 即完全负相关场合, 对应于 $A = \overline{B}$.

回归方程是

$$\mathbf{1}_B = P(B) + \frac{P(AB) - P(A)P(B)}{P(A)P(\overline{A})}(\mathbf{1}_A - P(A)) \quad (15)$$

$$\mathbf{1}_A = P(A) + \frac{P(AB) - P(A)P(B)}{P(B)P(\overline{B})}(\mathbf{1}_B - P(B)) \quad (16)$$

教学札记之二十四

母函数与分支过程

一、分支过程的定义

分支过程作为母体生长的一种概率模型最早由高尔顿 - 沃生进行研究, 它是应用随机过程中比较重要的一个, 是一种马尔可夫链, 初始用于描述家族姓氏的延续, 后来在核反应堆, 植物母体生长等领域也有应用.

由于它的本质是随机个独立同分布的整值随机变量之和, 因此很适合于用母函数来研究.

这个题材, 本身相当重要, 关于灭种概率的结论又十分深刻, 还能显示母函数法的巨大威力, 若非篇幅限制, 是可以写入正文的.

设每个质点能产生同类新质点, 一个质点作为第 0 代, 每个质点以概率 p_k $(k = 0, 1, 2, \cdots)$ 产生 k 个新质点, 各代间与各质点间

互不影响, 感兴趣的是第 n 代的质点数.

若以 X_n 记第 n 代的质点数. $X_0 \equiv 1$, X_1 有给定概率分布 $\{p_k,\ k=0,1,2,\cdots\}$, 而

$$X_n = y_1^{(n)} + y_2^{(n)} + \cdots + y_{X_{n-1}}^{(n)}, \quad n \geqslant 1 \tag{1}$$

其中 $y_i^{(n)}$ 对不同的 n 及不同的 i 都是相互独立、服从概率分布 $\{p_k\}$ 的随机变量, 即随机变量 X_n 是 X_{n-1} (也是随机变量) 个相互独立相同分布整值随机变量之和. 这样定义的随机过程 $\{X_n,\ n \geqslant 1\}$ 称为**分支过程**.

显然分支过程描述家族姓氏的延续, 只计男丁, 最关心家族是否能繁衍, 从而要探讨断绝的可能性. 在用来描述原子反应堆的中子数时, 关心的是能否形成核裂变.

这个模型显然很适合用《概率论基础》中母函数的方法来研究.

记概率分布 $\{p_k\}$ 的母函数为 $P(s)$, 即

$$P(s) = \sum_{k=0}^{\infty} p_k s^k \tag{2}$$

若以 $P_n(s)$ 记 X_n 的母函数, 则由该书 (4.4.14) 立刻得到

$$P_1(s) = P(s)$$

$$P_{n+1}(s) = P_n\big(P(s)\big) = P\big(P_n(s)\big), \quad n \geqslant 1 \tag{3}$$

二、灭种概率

为使问题的讨论有意义, 当然应假定 $0 < p_0 < 1$. 记 $\lambda_n = P\{X_n = 0\} = P_n(0)$, 它表示第 n 代时质点数为 0 的概率, 由定义知 λ_n 随 n 的增加而增加, 因此必有极限, 记为 λ, 称为该分支过程的**灭种概率**.

关于灭种概率有下列定理及基本引理.

定理 分支过程的灭种概率 λ 满足方程

$$\lambda = P(\lambda) \tag{4}$$

且是它的最小正根.

证明 事实上, 在 $0 < s < 1$ 时, $P(s)$ 是 s 的上升函数, 而 $\lambda_1 = P(0) = p_0 > 0$, 故

$$\lambda_2 = P_2(0) = P\big(P_1(0)\big) = P(\lambda_1) > P(0) = \lambda_1$$

于是用归纳法可证

$$\lambda_{n+1} = P(\lambda_n) > P(\lambda_{n-1}) = \lambda_n$$

因此序列 λ_n 单调上升地趋于一个极限 λ, 且 $0 < \lambda \leqslant 1$, 由于 $P(s)$ 在 $0 \leqslant s \leqslant 1$ 是连续的, 因此 λ 满足方程 (4).

如果 $u > 0$ 是方程 (4) 的任意一个根, 则 $\lambda_1 = P(0) < P(u) = u$, 因此用归纳法得知

$$\lambda_{n+1} = P(\lambda_n) < P(u) = u$$

这说明 $\lambda \leqslant u$, 因此 λ_n 趋于方程 $\lambda = P(\lambda)$ 的最小正根.

基本引理 设 $m = \sum_{k=0}^{\infty} kp_k$ 是后代期望数. 如果 $m \leqslant 1$, 则过程将在第 n 代之前灭种的概率趋于 1; 如果 $m > 1$, 则方程 $\lambda = P(\lambda)$ 存在唯一的根 $\lambda < 1$, 且 λ 是过程在有限多代后灭种概率的极限.

证明 由于 $\sum_{k=0}^{\infty} p_k = 1$, 故由魏尔斯特拉斯判别法知 $P(\lambda)$ 在 $[0,1]$ 中一致收敛. 显然对 $0 < \lambda \leqslant 1$, $P'(\lambda) > 0$, $P''(\lambda) \geqslant 0$, 所以 $P(\lambda)$ 是单调上升的下凸函数. 由假定 $P(0) = p_0 > 0$, $P(1) = \sum_{k=0}^{\infty} p_k = 1$, 故 $P(\lambda)$ 的图形只能如图 4–9 的 (a), (b), (c) 表示的那样, 它与对角线至多交于两点. 其中

$$m = \sum_{k=0}^{\infty} kp_k = P'(1)$$

当 $m \leqslant 1$ 时, $P(s)$ 在 $(1,1)$ 的斜率不大于 1, 于是方程 (4) 只存在一根 $\lambda = 1$; 而当 $m > 1$ 时, $P(s)$ 在 $(1,1)$ 的斜率大于 1, 于

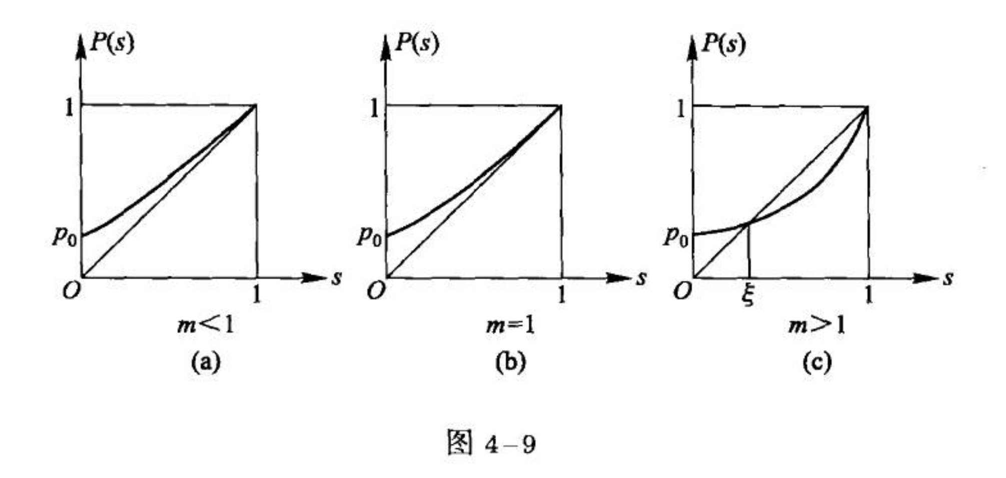


图 4-9

是方程 (4) 有一根 $0 < \xi < 1$, 此时 $\lambda = \xi$ 是灭种概率.

基本引理说明, 不但 "一代不如一代" 即 $m < 1$ 的家族必定灭种, 就是临界状态 $m = 1$ 俗称 "单子独孙" 的家族也必定灭种. 这在随机现象中是普适的, 因此这个结果常在概率论的各个分支中被引用, 故称基本引理.

第五章　极 限 定 理

章前引言

本章的结构依全书模块式结构的设计安排，分为三个层次，逐步深入.

首先，第一节讲概率论极限定理的策源地——伯努利试验，所用方法比较初等，它的结论与应用也比较直观，并且和频率与概率的主题密切相关. 作为其应用，第一、二章的一些计算问题在这里得到完满的解答.

其次，第三节讲独立同分布场合，特别适用于统计学. 第二节则为其作概念和工具上的准备，特别是证明了分布函数与特征函数对应的连续性，为特征函数成为研究概率极限定理的主力工具奠下基石.

最后，第四节讲强大数定律，第五节讲一般中心极限定理，都是很深刻的结果，证明的技巧性也很强.

课文导读

§ 5.1　伯努利试验场合的极限定理

在引进以随机变量与分布函数为代表的近代表述方法以及数字特征与特征函数等概念和工具之后，又回到伯努利试验. 这是早期概率论发展最丰饶的一块园地. 迄今为止，概率论最重要的理论

成果, 其源头几乎都可上溯至此; 即使在现代, 新的成果也常要拿到这里来做初步的验证.

奠定概率论学科基石的两大理论成果——大数定律与中心极限定理于 18 世纪初在这个园地相继诞生. 本节将以现代概率语言评述这两大经典结果.

由雅科布·伯努利明确表述并加以严格证明的第一个大数定律是概率论这一学科的灵魂和象征, 一百多年后被泊松推广并命名, 后被以切比雪夫为代表的俄罗斯学派发展成为很精细的理论.

对大数定律,《概率论基础》直接用 "矩法" 得出普遍结果, 再把上述两大经典定律作为特例, 最后论述了它的重要理论意义和实际应用价值.

1733 年棣莫弗找到 $p=\dfrac{1}{2}$ 时二项分布一般项的一个渐近表达式 (参看习题五 10 题), 概率论的标志——正态曲线正式登场, 半个多世纪后被拉普拉斯发展成为一个漂亮的极限定理, 之后关于它的研究占据概率论舞台近两百年, 最后获得中心极限定理的美名.

对中心极限定理,《概率论基础》在本节沿着棣莫弗的思路用斯特林公式作逼近, 导出局部极限定理, 然后用积分定义给出积分极限定理, 既采用当代表述形式, 又尽量保持历史发展的原生态.

最后是中心极限定理的应用, 该段包括重要内容:

(1) 导出大数定律. 说明中心极限定理比大数定律更强. 粗略讲, 大数定律表明 μ_n-np 是 $o(n)$, 只指出它是 n 的低阶无穷大, 而中心极限定理则进一步指出是 $O(\sqrt{n})$, 表明了 μ_n-np 是与 $\sqrt{n}$ 相同阶的无穷大.

(2) 提出用频率估计概率时的三类计算问题, 并给出实例. 实际上, 以后一般中心极限定理应用中遇到的也是同样的这三类问题, 不再细讲.

值得注意的还有: 置信区间估计这类统计方法在此以频率估计概率的例子形式首次出现.

(3) 完满解决了第二章提出的大量计算问题.

总之, 这是理论与应用并重的一节, 频率与概率的关系到这里基本讲清. 至此,《概率论基础》的"基础教程"目标实现.

§ 5.2 收敛性

上节使用矩法和直接展开证明了伯努利试验场合的大数定律与中心极限定理. 当推广到一般独立和这一更加复杂的场合, 这些方法常常施展不开, 必须另找新的工具. 经百年努力, 终于找到适合的工具特征函数. 在上章已引入特征函数, 讨论了它的性质, 并在研究多元正态分布中初显威力. 本章将用它来研究极限定理. 历史上它就是为此而引进的, 也因为它的使用而获得许多成果, 但是人们也发现, 为使整个证明过程严格, 有些基础必须夯实, 这些努力最后导致连续性定理的提出与证明, 它们归功于莱维与克拉美.

连续性定理证实了用特征函数列收敛来证明分布函数列收敛的合理性, 建立了这两种收敛的对应性. 为此必须搞清什么是分布函数列收敛, 研究表明适当的定义是弱收敛性, 而特征函数列则是内闭均敛, 连续性定理断言了这两种收敛的等价性.

第一小节引入分布函数列的弱收敛性, 并证明了海莱的几条定理, 它们在实函数论中有独立的兴趣, 但在这里被引用则是为下一小节证明正、逆极限定理 (合称连续性定理) 作准备. 依传统这些定理的证明细节全被写出, 因此本小节分析论证的分量很重, 任课教师可据具体情况取舍.

第二小节证明了正、逆极限定理, 其中逆极限定理的证明也是相当繁复的. 重要的是要明白正、逆极限定理的含意与用途.

第三小节定义了随机变量的四种收敛性并讨论了它们之间的关系, 这是将来研究各类概率极限定理的基础知识, 当然十分基本而重要. 应当说明, 对于以概率 1 收敛, 主要讨论放在第 4 节进行, 这里只出现定义.

这些收敛性的关系如图 5–1 所示:

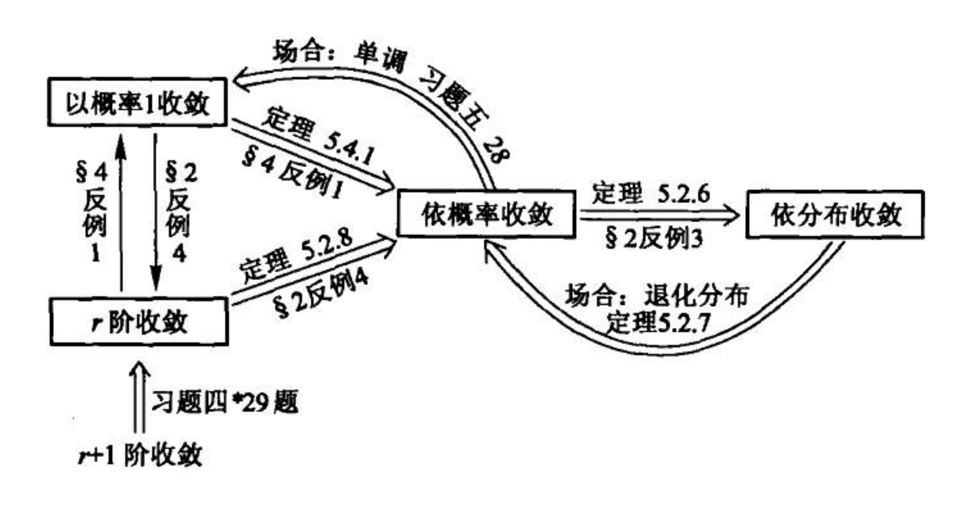


图 5－1

到此为止，已为极限定理的讨论做好一切准备.

第四小节给出了刻画特征函数充要条件的波赫纳尔－辛钦定理，表明非负定性是特征函数的关键性质.

第五小节顺便讨论了某种意义下帕斯卡分布向埃尔朗分布的收敛性，同理负二项分布将收敛于 Γ 分布.

§5.3 独立同分布场合的极限定理

本节把大数定律和中心极限定理推广到独立同分布的场合，特征函数显示了它的强大威力.

独立同分布可以说是概率论的第三个重要模型，另外两个是古典概型与伯努利概型，而且伯努利概型是它的简单特例.

当代统计学也把理论框架建立在独立同分布的基础上，即认为样本是总体的独立同分布观察即简单随机样本，这更增加了该模型的重要性. 因而本节的结果是数理统计中最常引用的.

本节证得的两个结果，在大数定律中辛钦大数定律，要求只是数学期望存在，这是十分自然的；在中心极限定理中林德伯格－莱维中心极限定理，成立的条件是 $0<\sigma^2<\infty$. 若 $\sigma^2=0$，则随机变量序列是常数序列，无须讨论极限定理. 若 $\sigma^2=\infty$，则无法规格

化, 中心极限定理也无从谈起. 因此这两个定理几乎没有条件, 是最终结果. 另一方面, 这两个定理（不要忘记伯努利大数定理与棣莫弗–拉普拉斯积分极限定理是其特例）在概率论中大约是引用频率最高的理论结果.

本节的另一看点是特征函数在处理独立和极限定理中的简单明了与巨大威力. 对照之下, 两个定理的证明过程极其相似, 而这也是用特征函数证明概率极限定理的通用程序, 值得读者再三玩味, 并通过若干习题演练, 掌握诀要, 以期将来在复杂场合能举一反三.

辛钦定理也可以用改良的矩法（对随机变量"截尾"使之有界）来证明, 但还是用特征函数法明快.

本来特征函数对应于分布函数只能用来处理依分布收敛, 但当收敛于常数时又等价于依概率收敛（这就是定理 5.2.7 的重要论断）, 因此也能用于处理大数定律. 这就大大增加了特征函数在处理概率极限定理中的重要性. 这点也是值得读者注意的.

《概率论基础》第三版中重点强调了这两个定理在数理统计学中的应用, 结合当代数理统计的理论框架, 读者将会有许多思考. 关于在蒙特卡罗法中的应用也如此. 参看教学札记之二十五"统计学和概率论".

多元中心极限定理引用极多, 但很少看到有教科书写出证明的. 《概率论基础》前有特征函数与多元正态的大量积累, 现有林德伯格–莱维中心极限定理, 顺理成章地写出了严密的证明.

根据模块式结构,《概率论基础》到本节为止, 构成"基本教程", 能满足一般学科对概率论的要求.

§ 5.4 强大数定律

本节讨论以概率 1 收敛及强大数定律. 博雷尔在正规数研究中证明了一条定理, 现在称之为博雷尔强大数定律, 提出了以概率 1 收敛性, 为后来的研究开辟了新道路.

前两个小节的内容可列入基本要求，其中博雷尔-康特立引理是理论研究中经常用到的，对理解以概率 1 收敛的涵义和证明强大数定律是基本的. 博雷尔大数定律针对伯努利试验场合，是伯努利大数定律的加强，而且这条定理的证明方法也相当初等，只要用到马尔可夫不等式及上述引理，相信大部分读者都能读懂.

到了独立同分布场合，只要 4 阶中心矩存在，则成立强大数定律（见习题五 46 题），证法与博雷尔强大数定律大体平行. 后来科尔莫戈罗夫得到最好的结果，这是第四小节的主题.

第一小节证明了以概率 1 收敛可推出依概率收敛，因此成立强大数定律必成立（弱）大数定律. 下面是成立弱大数定律而不成立强大数定律的一个例子:

若 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且

$$P\{X_1 = n\} = P\{X_1 = -n\} = \frac{C}{n^2 \log n}, \quad n = 3, 4, 5, \cdots$$

其中 C 是常数，等于 $\frac{1}{2}\Big(\sum\limits_{n=3}^{\infty} \frac{1}{n^2 \log n}\Big)^{-1}$.

这时 $E|X_1| = \infty$，故科尔莫戈罗夫强大数定律不成立，辛钦大数定律不能用，但能成立《概率论基础》(5.1.9)，即对 $\{X_n,\ n = 1, 2, \cdots\}$ 成立弱大数定律而不成立强大数定律.

第三小节是本节的要点与难点.

证明强大数定律主要使用两件工具，一是博雷尔-康特立引理，二是矩法的不等式. 马尔可夫不等式正如上述只能用于证明存在 4 阶矩的场合成立强大数定律，对更一般场合必须强化这些不等式，这个工作为科尔莫戈罗夫所完成，他证得的科尔莫戈罗夫不等式是该领域的主要武器. 《概率论基础》采用推广它为噶依克-瑞尼不等式的方式证明了一般场合的科尔莫戈罗夫强大数定律.

关于强大数定律的证明方法的进一步讨论见教学札记之二十九 “强大数定律证明途径评述”

本节是测度论概率论的重要组成部分，但本书还是严格坚持

其微积分基础假定.

§ 5.5 中心极限定理

长达 200 年历史的中心极限定理在 20 世纪有了突破性的进展. 首先是世纪初李雅普诺夫利用特征函数法证得了很一般的定理(《概率论基础》定理 5.5.4), 后来在 20 年代林德伯格找到很一般的充分条件——林德伯格条件, 最后在 30 年代费勒证明在某种条件 (后称费勒条件) 下, 林德伯格条件也是必要的. 这样一来, 中心极限定理最终获得解决. 本节介绍这些结果.

第一小节仔细地研究了林德伯格条件与费勒条件.

第二小节给出林德伯格 - 费勒定理的详细证明, 使用的工具是特征函数. 这个证明充分体现了特征函数法的精妙. 该证明过程的主体取材于费勒晚年的名著,《概率论基础》略加编排. 学习本节的要点之一是体会特征函数的精密使用.

第三小节作为上述定理的推论给出李雅普诺夫条件即《概率论基础》(5.5.34) 和另一个充分条件, 它们通常比林德伯格条件更便于验证.

本节的特征是分析性的, 即本节内容是分析概率论的重要组成部分, 可以避开测度论.

习题五的最后十道题目, 是本节的有机组成部分.

至此,《概率论基础》的 "高级教程" 全部讲完.

习题解答与评注

1. ξ 为非负随机变量, 若 $E\,\mathrm{e}^{a\xi}<\infty$, $(a>0)$, 则对任意 $x>0$,

$$P\{\xi\geqslant x\}\leqslant \mathrm{e}^{-ax}E\,\mathrm{e}^{a\xi}$$

证 记 ξ 的分布函数为 $F(x)$, 对任意 $x > 0$,

$$P\{\xi \geqslant x\} = \int_{y \geqslant x} \mathrm{d}F(y)$$

$$\leqslant \frac{1}{\mathrm{e}^{ax}} \int_{y \geqslant x} \mathrm{e}^{ay} \,\mathrm{d}F(y)$$

$$\leqslant \mathrm{e}^{-ax} \int_0^{\infty} \mathrm{e}^{ay} \,\mathrm{d}F(y)$$

$$= \mathrm{e}^{-ax} E\mathrm{e}^{a\xi}$$

【评注】又一种切比雪夫型不等式.

2. 若 $h(x) \geqslant 0$, ξ 为随机变量, 且 $Eh(\xi) < \infty$, 则关于任何 $C > 0$,

$$P\{h(\xi) \geqslant C\} \leqslant C^{-1} Eh(\xi)$$

证 记 $h(\xi)$ 的分布函数为 $F_h(x)$, 对任意 $C > 0$, 有

$$P\{h(\xi) \geqslant C\} = \int_{x \geqslant C} \mathrm{d}F_h(x)$$

$$\leqslant \frac{1}{C} \int_{x \geqslant C} x \mathrm{d}F_h(x)$$

$$\leqslant \frac{1}{C} \int_0^{\infty} x \,\mathrm{d}F_h(x)$$

$$= \frac{1}{C} Eh(\xi)$$

【评注】还是一种切比雪夫型不等式, 实质是马尔可夫不等式.

*3. (单边切比雪夫不等式) 设 ξ 为随机变量, $E\xi = 0$, $D\xi = \sigma^2 < \infty$, 则对任何一个 $a > 0$, 试证

$$P\{\xi \geqslant a\} \leqslant \frac{\sigma^2}{\sigma^2 + a^2}$$

证 记 ξ 的分布函数为 $F(x)$, 因为 $E\xi=0$, 所以

$$a=\int_{-\infty}^{\infty}(a-x)\,\mathrm{d}F(x)\leqslant\int_{-\infty}^{a}(a-x)\,\mathrm{d}F(x)$$
$$=\int_{-\infty}^{\infty}(a-x)J_a(x)\,\mathrm{d}F(x)$$

其中 $J_a(x)=\begin{cases}1, & x<a\\ 0, & x\geqslant a\end{cases}$, 由于 $a>0$, 两边平方得

$$a^2\leqslant\left(\int_{-\infty}^{\infty}(a-x)J_a(x)\mathrm{d}F(x)\right)^2$$
$$\overset{\text{柯西-施瓦兹不等式}}{\leqslant}\int_{-\infty}^{\infty}(a-x)^2\,\mathrm{d}F(x)\cdot\int_{-\infty}^{\infty}J_a^2(x)\,\mathrm{d}F(x)$$
$$=\int_{-\infty}^{\infty}(a-x)^2\,\mathrm{d}F(x)\cdot F(a)$$
$$=(a^2+\sigma^2)\,F(a)$$

所以

$$F(a)\geqslant\frac{a^2}{a^2+\sigma^2}$$

或等价地

$$P\{\xi\geqslant a\}=1-F(a)\leqslant1-\frac{a^2}{a^2+\sigma^2}=\frac{\sigma^2}{a^2+\sigma^2}$$

【评注】不再是一般推广而是有新意.

4. 设 $\{\xi_n,\ n\geqslant1\}$ 是独立随机变量序列, 对所有 $n\geqslant1$, $D\xi_n$ 存在且 $\dfrac{D\xi_n}{n}\longrightarrow0$, 试证: $\{\xi_n,\ n\geqslant1\}$ 服从大数定律.

证 由极限理论知, 对数列 $\{a_n\}$, 若 $a_n\longrightarrow a$, 可推得其前 n 项的算术平均构成的数列 $\left\{\dfrac{1}{n}\sum\limits_{k=1}^{n}a_k\right\}$ 有 $\dfrac{1}{n}\sum\limits_{k=1}^{n}a_k\longrightarrow a$.

本题已知 $\dfrac{D\xi_n}{n} \longrightarrow 0$, 则得

$$\frac{1}{n}\sum_{k=1}^{n}\frac{D\xi_k}{k} \longrightarrow 0, \quad n \to \infty$$

因而

$$0 \leqslant \frac{1}{n^2}D\Big(\sum_{k=1}^{n}\xi_k\Big) = \frac{1}{n}\sum_{k=1}^{n}\frac{D\xi_k}{n} \leqslant \frac{1}{n}\sum_{k=1}^{n}\frac{D\xi_k}{k} \longrightarrow 0, \quad n \to \infty$$

所以马尔可夫条件成立, 于是 $\{\xi_n, n \geqslant 1\}$ 服从大数定律.

【评注】这是很便于验证的一个充分条件, 完全可以写入教科书中.

5. 若 ξ_k 的分布列为

	$\sqrt{\ln k}$	$-\sqrt{\ln k}$
	$\frac{1}{2}$	$\frac{1}{2}$

试证大数定律适用于独立随机变量序列 $\{\xi_k\}$.

证

$$E\xi_k = 0$$

$$D\xi_k = \frac{1}{2}\left(\sqrt{\ln k}\right)^2 + \frac{1}{2}\left(-\sqrt{\ln k}\right)^2 = \ln k$$

$$\frac{1}{n^2}D\Big(\sum_{k=1}^{n}\xi_k\Big) = \frac{1}{n^2}\sum_{k=1}^{n}\ln k \leqslant \frac{1}{n^2}n\ln n = \frac{\ln n}{n} \longrightarrow 0,\ n \to \infty$$

满足马尔可夫条件, 故大数定律成立.

【评注】一道简单练习题. 马尔可夫条件是大数定律成立最普遍的一个充分条件, 所以首先试它. 当然证得习题 4 的结果后, 也可用它验证.

6. 验证概率分布如下给定的独立随机变量序列是否满足马尔可夫条件:

(1) $P\{X_k = \pm 2^k\} = \dfrac{1}{2}$;

(2) $P\{X_k=\pm 2^k\}=2^{-(2k+1)}$, $P\{X_k=0\}=1-2^{-2k}$;

(3) $P\{X_k=\pm k\}=\dfrac{1}{2}k^{-\frac{1}{2}}$, $P\{X_k=0\}=1-k^{-\frac{1}{2}}$.

提示: 几种基本类型, 如上题一样, 直接代入《概率论基础》(5.1.13) 验证.

答 (1) 不满足; (2) 满足; (3) 不满足.

【评注】几道典型题. 还将用于本章习题 47.

7. 若 ξ_k 具有有限方差, 服从同一分布, 但各 k 间, ξ_k 和 ξ_{k+1} 有相关, 而 ξ_k,ξ_l $(|k-l|\geqslant 2)$ 是独立的, 证明这时对 $\{\xi_k\}$ 大数定律成立.

证 由于 $|k-l|\geqslant 2$ 时, ξ_k 与 ξ_l 独立, 所以

$$\begin{aligned}\frac{1}{n^2}D\Big(\sum_{k=1}^n\xi_k\Big)&=\frac{1}{n^2}E\Big[\sum_{k=1}^n(\xi_k-E\xi_k)\Big]^2\\&=\frac{1}{n^2}\Big[\sum_{k=1}^nD\xi_k+2\sum_{k=1}^{n-1}\mathrm{cov}(\xi_k,\xi_{k+1})\Big]\\&=\frac{1}{n}D\xi_1+\frac{2}{n^2}\sum_{k=1}^{n-1}r_{k,k+1}D\xi_1\\&\leqslant\frac{D\xi_1}{n}+\frac{2}{n}D\xi_1\longrightarrow 0\quad(n\to\infty)\end{aligned}$$

马尔可夫条件成立, 所以对序列 $\{\xi_n\}$ 大数定律成立.

【评注】允许相邻变量相关, 仍能成立大数定律.

8. (伯恩斯坦定理) 已知随机变量序列 $\xi_1,\xi_2,\cdots$ 的方差有界: $D\xi_n\leqslant C$, 并且当 $|i-j|\to\infty$ 时, 相关系数 $r_{ij}\to 0$, 证明对 $\{\xi_k\}$ 成立大数定律.

证 因为 $|i-j|\to\infty$ 时, $r_{ij}\to 0$, 所以对任给 $\varepsilon>0$, 存在 N, 使当 $|i-j|>N$ 时有 $|r_{ij}|<\dfrac{\varepsilon}{C}$, 于是

$$\frac{1}{n^2}D\Big(\sum_{k=1}^n\xi_k\Big)=\frac{1}{n^2}\sum_{k=1}^nD\xi_k+\frac{2}{n^2}\sum_{1\leqslant i<j\leqslant n}r_{ij}\sigma_i\sigma_j$$

$$
\begin{aligned}
&\leqslant \frac{1}{n^2}nC+\frac{2}{n^2}\sum_{i<j}|r_{ij}|\sigma_i\sigma_j \\
&=\frac{C}{n}+\frac{2}{n^2}\sum_{0<j-i\leqslant N}|r_{ij}|\sigma_i\sigma_j+\frac{2}{n^2}\sum_{j-i>N}|r_{ij}|\sigma_i\sigma_j \\
&\leqslant \frac{C}{n}+\frac{2}{n^2}nNC+\frac{2}{n^2}\frac{n(n-1)}{2}\frac{\varepsilon}{C}C \\
&=\frac{C}{n}+\frac{2NC}{n}+\left(1-\frac{1}{n}\right)\varepsilon \\
&\leqslant \frac{C}{n}+\frac{2NC}{n}+\varepsilon
\end{aligned}
$$

(第二项 $\displaystyle\sum_{0<j-j\leqslant N}|r_{ij}|\sigma_i\sigma_j \leqslant \sum_{0<j-i\leqslant N} C$, 对每个固定的 i, 因 $i<j$, $j-i\leqslant N$, 所以最多对应 N 项, 而 i 可依次取 $1,2,\cdots,n$, 因而和式最多有 nN 项. 第三项中, 因 $i<j\leqslant n$, 所以对 i 分别取 $1,2,\cdots,n-1$ 时, 分别最多对应 $n-1,n-2,\cdots,1$ 项, 从而该和式最多含 $(n-1)+(n-2)+\cdots+1=\dfrac{n(n-1)}{2}$ 项.)

固定 N 后, 前两项在 $n\to\infty$ 时趋于 0, 最后一项可任意小, 故马尔可夫条件成立, 从而对 $\{\xi_i\}$ 成立大数定律.

【评注】伯恩斯坦 (1880—1968), 苏联概率统计学家, 对概率统计贡献甚多.

若方差有界, 只要不是前后变量一直保持相关, 大数定律仍能成立. 本题显然把上题结论大大往前推广.

从上两题可看出, 马尔可夫条件也能用于处理非独立序列的大数定律, 这是一大推进.

*9. (格涅坚科定理) 对随机变量序列 $\{\xi_i\}$, 若记 $\eta_n=\dfrac{1}{n}(\xi_1+\cdots+\xi_n)$, $a_n=\dfrac{1}{n}(E\xi_1+\cdots+E\xi_n)$, 则 $\{\xi_i\}$ 服从大数定律的充要条件是

$$
\lim_{n\to\infty}E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}=0
$$

证 充分性: $\dfrac{x^2}{1+x^2}$ 是 x 的增函数, 因此对 $\varepsilon > 0$, 有

$$
\begin{aligned}
P\{|\eta_n - a_n| \geqslant \varepsilon\} &= \int\limits_{|y-a_n|\geqslant\varepsilon} \mathrm{d}F_{\eta_n}(y) \\
&\leqslant \frac{1}{\dfrac{\varepsilon^2}{1+\varepsilon^2}} \int\limits_{|y-a_n|\geqslant\varepsilon} \frac{(y-a_n)^2}{1+(y-a_n)^2}\,\mathrm{d}F_{\eta_n}(y) \\
&\leqslant \frac{1+\varepsilon^2}{\varepsilon^2}\int_{-\infty}^{\infty} \frac{(y-a_n)^2}{1+(y-a_n)^2}\,\mathrm{d}F_{\eta_n}(y) \\
&= \frac{1+\varepsilon^2}{\varepsilon^2} E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}
\end{aligned}
$$

因而当 $E\left\{\dfrac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\} \to 0$ 时, 有 $P\{|\eta_n - a_n| \geqslant \varepsilon\} \to 0$, 于是 $\{\xi_i\}$ 服从大数定律.

必要性: 由于

$$
\begin{aligned}
0 &\leqslant E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\} \\
&= \int\limits_{|y-a_n|<\varepsilon} \frac{(y-a_n)^2}{1+(y-a_n)^2}\,\mathrm{d}F_{\eta_n}(y) + \int\limits_{|y-a_n|\geqslant\varepsilon} \frac{(y-a_n)^2}{1+(y-a_n)^2}\,\mathrm{d}F_{\eta_n}(y) \\
&\leqslant \frac{\varepsilon^2}{1+\varepsilon^2} P\{|\eta_n - a_n| < \varepsilon\} + 1\cdot P\{|\eta_n - a_n| \geqslant \varepsilon\} \\
&< \varepsilon^2 + P\{|\eta_n - a_n| \geqslant \varepsilon\}
\end{aligned}
$$

若 $\{\xi_i\}$ 服从大数定律, 由 ε 的任意性, 可得

$$
\lim_{n\to\infty} E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\} = 0
$$

【评注】有名的结果, 漂亮的证明.

顺便指出, 从证明中可以看出, 格涅坚科条件事实上是 $\eta_n - a_n \xrightarrow{P} 0$ 的充要条件, 不只用于大数定律.

10. 用斯特林公式证明: 当 $n\to\infty$, $m\to\infty$, $n-m\to\infty$, 而

$\dfrac{m}{n} \to 0$ 时,

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n} \sim \frac{1}{\sqrt{\pi n}}\,\mathrm{e}^{-m^2/n}$$

证 由斯特林公式

$$\begin{aligned}
\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n} &= \frac{(2n)!}{(n-m)!(n+m)!}\left(\frac{1}{2}\right)^{2n} \\
&= \frac{\sqrt{2\pi 2n}(2n)^{2n}\mathrm{e}^{-2n}\cdot \mathrm{e}^{\theta}}{\sqrt{2\pi(n-m)}(n-m)^{n-m}\mathrm{e}^{-(n-m)}} \\
&\quad \cdot \frac{1}{\sqrt{2\pi(n+m)}(n+m)^{n+m}\mathrm{e}^{-(n+m)}}\left(\frac{1}{2}\right)^{2n} \\
&= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2n}{n^2-m^2}}\left(\frac{n}{n-m}\right)^{n-m}\left(\frac{n}{n+m}\right)^{n+m}\mathrm{e}^{\theta} \\
&= \frac{1}{\sqrt{n\pi}}\left(1-\frac{m^2}{n^2}\right)^{-\frac{1}{2}}\left(1-\frac{m}{n}\right)^{-(n-m)}\left(1+\frac{m}{n}\right)^{-(n+m)}\mathrm{e}^{\theta}
\end{aligned}$$

其中 $|\theta| < \dfrac{1}{12}\left(\dfrac{1}{2n}+\dfrac{1}{n-m}+\dfrac{1}{n+m}\right)$.

由 $\ln(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \cdots$ 得

$$\begin{aligned}
&\ln\left[\left(1-\frac{m^2}{n^2}\right)^{-\frac{1}{2}}\left(1-\frac{m}{n}\right)^{m-n}\left(1+\frac{m}{n}\right)^{-(n+m)}\mathrm{e}^{\theta}\right] \\
&= \frac{1}{2}\left(\frac{m}{n}\right)^2 + (m-n)\left[-\frac{m}{n} - \frac{1}{2}\left(\frac{m}{n}\right)^2 - \frac{1}{3}\left(\frac{m}{n}\right)^3\right] \\
&\quad - (n+m)\left[\frac{m}{n} - \frac{1}{2}\left(\frac{m}{n}\right)^2 + \frac{1}{3}\left(\frac{m}{n}\right)^3\right] + \theta + \cdots \\
&= -\frac{m^2}{n} - \frac{2}{3}\frac{m^4}{n^3} + \frac{1}{2}\frac{m^2}{n^2} + \theta + \cdots
\end{aligned}$$

而 $\dfrac{m^4}{n^3} = \dfrac{m^2}{n}\dfrac{m^2}{n^2}$, 假定 $\dfrac{m^2}{n}$ 是一有界量, 则 $\dfrac{m^4}{n^3} = o\left(\dfrac{m}{n}\right)$, 故有

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n} \sim \frac{1}{\sqrt{\pi n}}\,\mathrm{e}^{-m^2/n}$$

【评注】这是 1733 年棣莫弗论文的现代表述, 以正态曲线逼

近 $p=\frac{1}{2}$ 时的二项分布, 正是这一工作使棣莫弗的名字永载概率史册.

11. 用 (5.1.27) 计算 $b(5;500,0.01)$ 及 $b(40;10\,000,0.005)$ 并与精确值比较.

提示: 取前两项.

答 (1) $b(5;500,0.01)$, 用 (5.1.27) 计算为 0.179 31, 精确值为 0.176 35.

(2) $b(40;10\,000,0.005)$, 用 (5.1.27) 计算为 0.021 4, 精确值为 0.021 4.

【评注】比起用 (5.1.38) 计算, 一般能得出更精确的结果.

现代用一只小型科学计算器几秒钟就能算出二项分布的精确值, 但正是当年计算的困难才推动人类发现正态分布律!

12. 某计算机系统有 120 个终端, 每个终端有 5% 时间在使用, 若各个终端使用与否是相互独立的, 试求有 10 个或更多终端在使用的概率.

提示: 用修正公式, 见《概率论基础》(5.1.41).

答 0.072.

【评注】代表一类非常实用的实际问题.

*13. 求证, 在 $x>0$ 时, 有不等式

$$\frac{x}{1+x^2}\mathrm{e}^{-\frac{x^2}{2}}\leqslant\int_x^\infty \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\leqslant\frac{1}{x}\mathrm{e}^{-\frac{x^2}{2}}$$

证 当 $x>0$, 有

$$\int_x^\infty \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\leqslant\int_x^\infty \frac{t}{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\frac{1}{x}\left(-\mathrm{e}^{-\frac{t^2}{2}}\right)\Big|_x^\infty=\frac{1}{x}\mathrm{e}^{-\frac{x^2}{2}}$$

另一方面

$$\int_x^\infty \mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\geqslant\int_x^\infty \frac{t^4+2t^2-1}{t^4+2t^2+1}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$

$$= \int_x^{\infty} \mathrm{d}\left(-\frac{t}{1+t^2}\mathrm{e}^{-\frac{t^2}{2}}\right) = -\frac{t}{1+t^2}\mathrm{e}^{-\frac{t^2}{2}}\Big|_x^{\infty}$$

$$= \frac{x}{1+x^2}\mathrm{e}^{-\frac{x^2}{2}}$$

【评注】很有用的不等式. 当 x 比较大, 在正态分布数值表上找不到时就要用这个不等式估算.

14. 用棣莫弗-拉普拉斯定理证明, 在伯努利试验中, 若 $0 < p < 1$, 则不管 K 是如何大的常数, 总有

$$P\{|\mu_n - np| < K\} \to 0 \quad (n \to \infty)$$

证 对任意固定 K, 对任意 $\delta > 0$, 存在 N_1, 有

$$\int_{-\frac{K}{\sqrt{N_1 pq}}}^{\frac{K}{\sqrt{N_1 pq}}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x < \frac{\delta}{2}$$

又由棣莫弗-拉普拉斯积分定理, 有

$$P\left\{\frac{|\mu_n - np|}{\sqrt{npq}} < \frac{K}{\sqrt{N_1 pq}}\right\} \longrightarrow \int_{-\frac{K}{\sqrt{N_1 pq}}}^{\frac{K}{\sqrt{N_1 pq}}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x$$

所以存在 N_2, 当 $n > N_2$ 时, 有

$$P\left\{\frac{|\mu_n - np|}{\sqrt{npq}} < \frac{K}{\sqrt{N_1 pq}}\right\} < \int_{-\frac{K}{\sqrt{N_1 pq}}}^{\frac{K}{\sqrt{N_1 pq}}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x + \frac{\delta}{2} < \delta$$

取 $N = \max(N_1, N_2)$, 当 $n > N$ 时,

$$P\{|\mu_n - np| < K\} \leqslant P\left\{\frac{|\mu - np|}{\sqrt{npq}} < \frac{K}{\sqrt{N_1 pq}}\right\} < \delta$$

从而有

$$P\{|\mu_n - np| < K\} \to 0 \quad (n \to \infty)$$

【评注】青山遮不住, 毕竟东流去. 偏差 $|\mu_n - np|$ 的"阶数"是 $\sqrt{npq}$.

15. 用切比雪夫不等式确定当掷一均匀铜币时, 需投多少次才能保证使得正面出现的概率在 0.4 至 0.6 之间的概率不小于 90%,

并用正态逼近计算同一问题.

提示: 前者直接用切比雪夫不等式计算, 这是该不等式的用途之一, 值得一做. 后者用《概率论基础》(5.1.30), 属频率估计概率中第二类问题.

答 用切比雪夫不等式, $n \geqslant 250$ 次;

用正态逼近, $n \geqslant 68$ 次.

【评注】二者差距太大, 你是如何想的?

16. 用切比雪夫不等式及棣莫弗–拉普拉斯极限定理估计下面概率:

$$P\left\{\left|\frac{\mu_n}{n}-p\right| \geqslant \varepsilon\right\}$$

并进行比较, 这里 μ_n 是 n 次伯努利试验中成功总次数, p 为每次成功的概率.

解 用切比雪夫不等式,

$$P\left\{\left|\frac{\mu_n}{n}-p\right| \geqslant \varepsilon\right\} \leqslant \frac{npq}{\varepsilon^2 n^2}=\frac{pq}{n\varepsilon^2}$$

用极限定理,

$$\begin{aligned} P\left\{\left|\frac{\mu_n}{n}-p\right| \geqslant \varepsilon\right\} &= P\left\{\frac{|\mu_n-np|}{\sqrt{npq}} \geqslant \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right\} \\ &\approx \frac{2}{\sqrt{2\pi}}\int_{\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}}^{\infty} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t \\ &< \sqrt{\frac{2}{\pi}}\,\frac{\sqrt{pq}}{\varepsilon\sqrt{n}}\int_{\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}}^{\infty} t\,\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t \\ &= \frac{1}{\varepsilon}\sqrt{\frac{2pq}{n\pi}}\,\mathrm{e}^{-\frac{\varepsilon^2 n}{2pq}} = o\left(\frac{pq}{n\varepsilon^2}\right) \quad (n \to \infty) \end{aligned}$$

所以当 n 较大时, 极限定理能得到更精确的估计.

【评注】切比雪夫不等式只要知道均值与方差, 不管分布如何, 都可对概率作估算, 应用范围很广, 但有一利必有一弊, 它的估计就不可能很准确. 极限定理涉及精确分布, 只要 n 大到它成立, 得

出的结论就较准确.

17. 现有一大批种子, 其中良种占 1/6, 今在其中任选 6 000 粒, 试问在这些种子中良种所占的比例与 1/6 之差小于 1% 的概率是多少?

提示: 积分极限定理的直接应用《概率论基础》(5.1.29).

答 0.96.

【评注】用积分极限定理计算频率估计概率的第一类问题: 已知 n, p, ε 求概率 β.

18. 种子中良种占 1/6, 我们有 99% 的把握断定在 6 000 粒种子中良种所占的比例与 1/6 之差是多少? 这时相应的良种粒数落在哪个范围内.

提示: 用《概率论基础》(5.1.31).

答 差是 0.012 4. 良种粒数落在 925 粒与 1 075 粒之间.

【评注】同上, 属第三类问题, 求 ε.

19. 若飞机乘客购票后按期搭机的概率为 p, 各乘客的行动假定是独立的, 试问一架 200 座飞机售出 202 张机票不发生超座的概率. 对 $p = 0.97, 0.96, 0.95$, 计算上述概率.

解 记 μ 为购票后按期搭机的旅客数, 则

$$P\{\mu \leqslant 200\} = P\left\{ \frac{\mu - 202p}{\sqrt{202p(1-p)}} \leqslant \frac{200 - 202p}{\sqrt{202p(1-p)}} \right\}$$

$$\approx \Phi\left(\frac{200 + 0.5 - 202p}{\sqrt{202p(1-p)}} \right)$$

p	0.97	0.96	0.95
P	0.970	0.991	0.997

【评注】从另一个角度讨论超售问题: 固定超售数额, 对不同假定的 p 值计算不出现拒登机的概率.

*20. 设分布函数列 $\{F_n(x)\}$ 弱收敛于连续的分布函数 $F(x)$, 试证这收敛对 $x \in \mathbf{R}^1$ 是一致的.

证 对实数 $M < N$, 有

$$\sup_{x\in\mathbf{R}^1}\left|F_n(x)-F(x)\right| \leqslant I_{n,1}+I_{n,2}+I_{n,3}+I_{n,4}+I_{n,5} \qquad (*)$$

其中

$$I_{n,1}=\sup_{-\infty<x\leqslant M}F_n(x)=F_n(M)$$

$$I_{n,2}=\sup_{-\infty<x\leqslant M}F(x)=F(M)$$

$$I_{n,3}=\sup_{x\in[M,N]}\left|F_n(x)-F(x)\right|$$

$$I_{n,4}=\sup_{N<x<\infty}\left[1-F(x)\right]=1-F(N)$$

$$I_{n,5}=\sup_{N<x<\infty}\left[1-F_n(x)\right]=1-F_n(N)$$

对任给 $\varepsilon>0$, 取 M 和 N 分别使 $I_{n,2}<\varepsilon$ 和 $I_{n,4}<\varepsilon$. 由于 $F_n \xrightarrow{w} F$ 且 F 在 $\mathbf{R}^1$ 上连续, 故又有 $\lim\limits_{n\to\infty}I_{n,1}<\varepsilon$, $\lim\limits_{n\to\infty}I_{n,5}<\varepsilon$ 和 $\lim\limits_{n\to\infty}I_{n,3}=0$. 于是在 $(*)$ 中先令 $n\to\infty$, 再令 $\varepsilon\to 0$ 即得所要结论.

【评注】这是很重要的一个结论, 省去极限定理讨论中的许多麻烦.

*21. 设 $\{F_n(x)\}$ 为一列正态分布函数, 收敛于分布函数 $F(x)$, 试证 $F(x)$ 也是正态分布函数.

证 设 $F_n(x)$, $F(x)$ 的特征函数分别为 $f_n(t)$ 和 $f(t)$, 由题意

$$\lim_{n\to\infty}f_n(t)=\lim_{n\to\infty}\mathrm{e}^{\mathrm{i}\mu_n t-\frac{1}{2}\sigma_n^2t^2}=f(t)$$

因而有

$$\left|f_n(t)\right|=\mathrm{e}^{-\frac{1}{2}\sigma_n^2t^2}\longrightarrow\left|f(t)\right|$$

因特征函数 $f(t)$ 的连续性且 $f(0)=1$, 故存在 $t_0\neq 0$ 使 $f(t_0)\neq 0$, 故而

$$-\frac{1}{2}\sigma_n^2t_0^2\longrightarrow\log\left|f(t_0)\right|\neq-\infty$$

于是得 $\lim\limits_{n\to\infty}\sigma_n^2$ 存在, 记该极限为 σ^2.

这样就有

$$\mathrm{e}^{\mathrm{i}\mu_n t}=f_n(t)\,\mathrm{e}^{\frac{1}{2}\sigma_n^2t^2}\longrightarrow f(t)\mathrm{e}^{\frac{1}{2}\sigma^2t^2}$$

且有

$$\left|f(t)\mathrm{e}^{\frac{1}{2}\sigma^2t^2}\right|=\lim_{n\to\infty}\left|\mathrm{e}^{i\mu_n t}\right|=1$$

故同样可得 $\lim\limits_{n\to\infty}\mu_n$ 存在, 记 $\lim\limits_{n\to\infty}\mu_n=\mu$, 于是

$$\mathrm{e}^{\mathrm{i}\mu_n t-\frac{1}{2}\sigma_n^2t^2}=f_n(t)\longrightarrow f(t)=\mathrm{e}^{\mathrm{i}\mu\, t-\frac{1}{2}\sigma^2t^2}$$

所以 $F(x)$ 也是正态分布.

【评注】正态分布的重要性质之一. 证明过程再次显示特征函数的威力. 不禁要问, 如果没有特征函数, 要怎样证明这个结果?!

*22. 试证若正态随机变量序列依概率收敛, 则其数学期望与方差也收敛.

证 若以 $F_n(x)$ 及 $F(x)$ 分别记正态变量 X_n 及 X 的分布函数, 由 $X_n\xrightarrow{P}X$ 可推得 $X_n\xrightarrow{L}X$, 从而 $F_n(x)\xrightarrow{w}F(x)$, 由本章习题 21 则可得结论.

【评注】也是一个很有用的性质.

*23. 若 $\boldsymbol{X}_n$ 为多维正态随机向量, $\boldsymbol{X}_n\xrightarrow{P}\boldsymbol{X}$, 试证 $\boldsymbol{X}$ 为正态向量.

证 首先注意教科书上的一个定理: $\boldsymbol{\xi}=(\xi_1,\cdots,\xi_k)^{\mathrm{T}}$ 服从 k 元正态分布 $N(\boldsymbol{a},\boldsymbol{\Sigma})$ 的充要条件是它的任何一个线性组合 $\zeta=\sum\limits_{j=1}^{k}l_j\xi_j$ 服从一元正态分布 $N\Big(\sum\limits_{j=1}^{k}l_ja_j,\ \sum\limits_{i,j=1}^{k}l_il_j\sigma_{ij}\Big)$.

设 $\boldsymbol{X_n}=\big(X_{n,1},\cdots,X_{n,k}\big)^{\mathrm{T}}$, $\boldsymbol{X}=\big(X_1,\cdots,X_k\big)^{\mathrm{T}}$, 由 $\boldsymbol{X}_n\xrightarrow{P}\boldsymbol{X}$, 得任一线性组合 $Y_n=\sum\limits_{j=1}^{k}l_jX_{n,j}\xrightarrow{P}Y=\sum\limits_{j=1}^{k}l_jX_j$, 因此有 $Y_n\xrightarrow{L}Y$, $F_{Y_n}(x)\xrightarrow{w}F_Y(x)$, 由前述定理知 Y_n 服从一元正态分

布, 再由本章习题 21 知 Y 也服从一元正态分布, 再由前述定理得 $\boldsymbol{X}$ 为多元正态向量.

【评注】把多维正态的证明问题化为一维问题, 切记!本章习题 *26 也立一功.

24. 若 X_n 的概率分布为

	0	n
	$1-\frac{1}{n}$	$\frac{1}{n}$

试证相应的分布函数列收敛, 但矩不收敛.

证

$$F_n(x)=P\{X_n<x\}=\begin{cases}0 & ,\quad 当\ x\leqslant 0\\ 1-\dfrac{1}{n}, & \quad 当\ 0<x\leqslant n\\ 1 & ,\quad 当\ x>n\end{cases}$$

令 $n\to\infty$, 得

$$F_n(x)\longrightarrow F(x)=\begin{cases}0, & x\leqslant 0\\ 1, & x>0\end{cases}$$

这表明分布函数列收敛.

但 $EX_n=1$, $EX_n \not\to 0 \quad (n\to\infty)$.

当 $k>1$ 时,

$$EX_n^k=n^k\cdot\frac{1}{n}=n^{k-1}$$

$$E(X_n-EX_n)^k=E(X_n-1)^k=(-1)^k\left(1-\frac{1}{n}\right)+(n-1)^k\cdot\frac{1}{n}$$

所以当 $n\to\infty$ 时, $EX_n^k\to\infty$, $E(X_n-EX_n)^k\to\infty$, 由此知 X_n 的原点矩、中心矩均不收敛.

【评注】很容易想出的一个反例, 第五章 §2 例 4 亦可参照.

*25. (斯卢茨基) 随机变量序列 $\{\xi_n\}$ 具有分布函数列 $\{F_n(x)\}$, 且 $F_n(x)\to F(x)$, 又 $\{\eta_n\}$ 依概率收敛于常数 $C>0$, 试证: (1) $\zeta_n=$

$\xi_n+\eta_n$ 的分布函数收敛于 $F(x-C)$; (2) $\zeta_n=\dfrac{\xi_n}{\eta_n}$ 的分布函数收敛于 $F(Cx)$.

证 (1) 对任意 $\varepsilon>0$, 一方面

$$
\begin{aligned}
&P\{\xi_n+\eta_n<x\}\\
&\quad=P\{\xi_n+\eta_n<x,\ |\eta_n-C|<\varepsilon\}+P\{\xi_n+\eta_n<x,\ |\eta_n-C|\geqslant\varepsilon\}\\
&\quad\leqslant P\{\xi_n+\eta_n<x,\ |\eta_n-C|<\varepsilon\}+P\{|\eta_n-C|\geqslant\varepsilon\}\\
&\quad\leqslant P\{\xi_n<x-(C-\varepsilon)\}+P\{|\eta_n-C|\geqslant\varepsilon\}
\end{aligned}
$$

而 $P\{|\eta_n-C|\geqslant\varepsilon\}\longrightarrow 0\ (n\to\infty)$, 所以有

$$\overline{\lim_{n\to\infty}}\, F_{\zeta_n}(x)\leqslant\overline{\lim_{n\to\infty}}\, F_n(x-C+\varepsilon)$$

另一方面

$$
\begin{aligned}
&P\{\xi_n<x-(C+\varepsilon)\}\\
&\quad=P\{\xi_n<x-(C+\varepsilon),\ |\eta_n-C|<\varepsilon\}\\
&\qquad+P\{\xi_n<x-(C+\varepsilon),\ |\eta_n-C|\geqslant\varepsilon\}\\
&\quad\leqslant P\{\xi_n<x-(C+\varepsilon),\ |\eta_n-C|<\varepsilon\}+P\{\eta_n-C|\geqslant\varepsilon\}\\
&\quad\leqslant P\{\xi_n+\eta_n<x\}+P\{|\eta_n-C|\geqslant\varepsilon\}
\end{aligned}
$$

还是因为 $P\{|\eta_n-C|\geqslant\varepsilon\}\longrightarrow 0\ (n\to\infty)$, 所以又有

$$\underline{\lim_{n\to\infty}}\, F_n(x-C-\varepsilon)\leqslant\underline{\lim_{n\to\infty}}\, F_{\zeta_n}(x)$$

由于分布函数至多有可列个不连续点, 因此可取 $x-C$ 及 $x-C\pm\varepsilon$ 为 $F(x)$ 的连续点, 且将上两式合并考虑之, 则得

$$\underline{\lim_{n\to\infty}}\, F_n(x-C-\varepsilon)\leqslant\underline{\lim_{n\to\infty}}\, F_{\zeta_n}(x)\leqslant\overline{\lim_{n\to\infty}}\, F_{\zeta_n}(x)\leqslant\overline{\lim_{n\to\infty}}\, F_n(x-C+\varepsilon)$$

再由 $F_n(x)\longrightarrow F(x)$ 可知

$$F(x-C-\varepsilon)\leqslant\underline{\lim_{n\to\infty}}\, F_{\zeta_n}(x)\leqslant\overline{\lim_{n\to\infty}}\, F_{\zeta_n}(x)\leqslant F(x-C+\varepsilon)$$

最后令 $\varepsilon \to 0$, 即得

$$\lim_{n\to\infty} F_{\zeta_n}(x) = F(x-C)$$

(2) 对任给 $0<\varepsilon<C$,

$$\begin{aligned}
&P\left\{\frac{\xi_n}{\eta_n}<x\right\}\\
&=P\left\{\frac{\xi_n}{\eta_n}<x,\ |\eta_n-C|<\varepsilon\right\}+P\left\{\frac{\xi_n}{\eta_n}<x,\ |\eta_n-C|\geqslant\varepsilon\right\}\\
&\leqslant P\left\{\frac{\xi_n}{\eta_n}<x,\ |\eta_n-C|<\varepsilon\right\}+P\{|\eta_n-C|\geqslant\varepsilon\}\\
&\leqslant P\{\xi_n<(C+\varepsilon)x\}+P\{|\eta_n-C|\geqslant\varepsilon\}
\end{aligned}$$

故有

$$\overline{\lim_{n\to\infty}} F_{\zeta_n}(x) \leqslant \overline{\lim_{n\to\infty}} F_n\big((C+\varepsilon)x\big)$$

另一方面,

$$\begin{aligned}
&P\{\xi_n<(C-\varepsilon)x\}\\
&=P\{\xi_n<(C-\varepsilon)x,\ |\eta_n-C|<\varepsilon\}+P\{\xi_n<(C-\varepsilon)x,\ |\eta_n-C|\geqslant\varepsilon\}\\
&\leqslant P\{\xi_n<(C-\varepsilon)x,\ |\eta_n-C|<\varepsilon\}+P\{|\eta_n-C|\geqslant\varepsilon\}\\
&\leqslant P\left\{\frac{\xi_n}{\eta_n}<x\right\}+P\{|\eta_n-C|\geqslant\varepsilon\}
\end{aligned}$$

得

$$\underline{\lim_{n\to\infty}} F_n\big((C-\varepsilon)x\big) \leqslant \underline{\lim_{n\to\infty}} F_{\zeta_n}(x)$$

可取 Cx 及 $(C\pm\varepsilon)x$ 为 $F(x)$ 的连续点, 于是,

$$F\big((C-\varepsilon)x\big) \leqslant \underline{\lim_{n\to\infty}} F_{\zeta_n}(x) \leqslant \overline{\lim_{n\to\infty}} F_{\zeta_n}(x) \leqslant F\big((C+\varepsilon)x\big)$$

令 $\varepsilon \to 0$ 得

$$\lim_{n\to\infty} F_{\zeta_n}(x) = F(Cx)$$

【评注】本题与以下两道习题分别涉及随机变量收敛性的三类问题, 结论常用, 证法典型.

苏联数学家斯卢茨基在 1925 年证得: 若一个序列依分布收敛, 一个序列依概率收敛, 则其和、差、积、商必依分布收敛. 这里举其要为题.

接下来很自然会提问: 若两个序列都是依分布收敛, 其和是否依分布收敛? 这个结论在何种条件下成立?

*26. 试证:

(1) $X_n \xrightarrow{P} X \Rightarrow X_n - X \xrightarrow{P} 0$;

(2) $X_n \xrightarrow{P} X,\ X_n \xrightarrow{P} Y \Rightarrow P\{X = Y\} = 1$;

(3) $X_n \xrightarrow{P} X \Rightarrow X_n - X_m \xrightarrow{P} 0\ (n,\ m \to \infty)$;

(4) $X_n \xrightarrow{P} X,\ Y_n \xrightarrow{P} Y \Rightarrow X_n \pm Y_n \xrightarrow{P} X \pm Y$;

(5) $X_n \xrightarrow{P} X$, k 是常数 $\Rightarrow kX_n \xrightarrow{P} kX$;

(6) $X_n \xrightarrow{P} X \Rightarrow X_n^2 \xrightarrow{P} X^2$;

(7) $X_n \xrightarrow{P} a,\ Y_n \xrightarrow{P} b$, $a,\ b$ 是常数 $\Rightarrow X_nY_n \xrightarrow{P} ab$;

(8) $X_n \xrightarrow{P} 1 \Rightarrow X_n^{-1} \xrightarrow{P} 1$;

(9) $X_n \xrightarrow{P} a,\ Y_n \xrightarrow{P} b$, a, b 是常数, $b \neq 0 \Rightarrow X_nY_n^{-1} \xrightarrow{P} ab^{-1}$;

(10) $X_n \xrightarrow{P} X$, Y 是随机变量 $\Rightarrow X_nY \xrightarrow{P} XY$;

(11) $X_n \xrightarrow{P} X,\ Y_n \xrightarrow{P} Y \Rightarrow X_nY_n \xrightarrow{P} XY$.

证 对任给 $\varepsilon > 0$,

(1) $P\{|X_n - X - 0| \geqslant \varepsilon\} = P\{|X_n - X| \geqslant \varepsilon\} \longrightarrow 0$

所以

$$X_n - X \xrightarrow{P} 0 \quad (n \to \infty)$$

(2)
$$\begin{aligned} P\{|X - Y| \geqslant \varepsilon\} &= P\{|X - X_n + X_n - Y| \geqslant \varepsilon\} \\ &\leqslant P\left\{|X_n - X| \geqslant \frac{\varepsilon}{2}\right\} + P\left\{|X_n - Y| \geqslant \frac{\varepsilon}{2}\right\} \\ &\longrightarrow 0 \quad (n \to \infty) \end{aligned}$$

由 ε 的任意性得 $P\{X \neq Y\} = 0$, 所以 $P\{X = Y\} = 1$.

(3) $P\{|X_n - X_m| \geqslant \varepsilon\} = P\{|X_n - X + X - X_m| \geqslant \varepsilon\}$

$$\leqslant P\left\{|X_n - X| \geqslant \frac{\varepsilon}{2}\right\} + P\left\{|X_m - X| \geqslant \frac{\varepsilon}{2}\right\}$$

$$\longrightarrow 0 \quad (n,\ m \to \infty)$$

所以 $X_n - X_m \xrightarrow{P} 0\ (n,\ m \to \infty)$.

(4) $P\{|X_n \pm Y_n - (X \pm Y)| \geqslant \varepsilon\}$

$$= P\{|(X_n - X) \pm (Y_n - Y)| \geqslant \varepsilon\}$$

$$\leqslant P\left\{|X_n - X| \geqslant \frac{\varepsilon}{2}\right\} + P\left\{|Y_n - Y| \geqslant \frac{\varepsilon}{2}\right\}$$

$$\longrightarrow 0 \quad (n \to \infty)$$

所以得 $X_n \pm Y_n \xrightarrow{P} X \pm Y$.

(5) $k = 0$, 显然成立. 若 $k \neq 0$,

$$P\{|kX_n - kX| \geqslant \varepsilon\} = P\{|k||X_n - X| \geqslant \varepsilon\}$$

$$= P\left\{|X_n - X| \geqslant \frac{\varepsilon}{|k|}\right\}$$

$$\longrightarrow 0 \quad (n \to \infty)$$

所以 $kX_n \xrightarrow{P} kX$.

(6) 对任给 $\delta > 0$, 存在 $M > 0$ 及正整数 N, 使当 $n > N$ 时有

$$P\left\{|X| \geqslant \frac{M}{2}\right\} < \frac{\delta}{4}$$

$$P\left\{|X_n - X| \geqslant \frac{M}{2}\right\} < \frac{\delta}{4}$$

$$P\left\{|X_n - X| \geqslant \frac{\varepsilon}{M}\right\} < \frac{\delta}{2}$$

一方面,

$P\{|X_n| \geqslant M\}$

$$= P\left\{|X_n| \geqslant M,\ |X| < \frac{M}{2}\right\} + P\left\{|X_n| \geqslant M,\ |X| \geqslant \frac{M}{2}\right\}$$

$$\leqslant P\left\{|X_n - X| \geqslant \frac{M}{2}\right\} + P\left\{|X| \geqslant \frac{M}{2}\right\}$$

$$< \frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2}$$

另一方面,

$$P\{|X_n||X_n - X| \geqslant \varepsilon\}$$

$$= P\{|X_n||X_n - X| \geqslant \varepsilon,\ |X_n| = 0\}$$

$$+ P\{|X_n||X_n - X| \geqslant \varepsilon,\ 0 < |X_n| < M\}$$

$$+ P\{|X_n||X_n - X| \geqslant \varepsilon,\ |X_n| \geqslant M\}$$

$$\leqslant P\left\{|X_n - X| \geqslant \frac{\varepsilon}{M}\right\} + P\{|X_n| \geqslant M\}$$

$$< \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

故 $X_n^2 - X_n X \xrightarrow{P} 0.$

同理可证 $X_n X - X^2 \xrightarrow{P} 0.$

因此由 (4) 知 $X_n^2 - X^2 = X_n^2 - X_n X + X_n X - X^2 \xrightarrow{P} 0.$

(7) 由 (4),

$$X_n + Y_n \xrightarrow{P} a + b$$

再由 (6),

$$(X_n + Y_n)^2 \xrightarrow{P} (a+b)^2, \quad X_n^2 \xrightarrow{P} a^2, \quad Y_n^2 \xrightarrow{P} b^2$$

故知 $X_n Y_n \xrightarrow{P} ab.$

(8) $P\{|X_n^{-1} - 1| \geqslant \varepsilon\}$

$$= P\left\{\frac{|1 - X_n|}{|X_n|} \geqslant \varepsilon\right\}$$

$$\leqslant P\left\{\frac{|X_n - 1|}{|X_n|} \geqslant \varepsilon,\ |X_n| > \frac{1}{2}\right\} + P\left\{|X_n| \leqslant \frac{1}{2}\right\}$$

$$\leqslant P\left\{|X_n-1|>\frac{\varepsilon}{2}\right\}+P\left\{|X_n-1|\geqslant\frac{1}{2}\right\}$$

$$\longrightarrow 0 \quad (n\to\infty)$$

所以 $X_n^{-1}\xrightarrow{P}1$.

(9) 利用本题 (5), (7), (8) 的结论.

(10) 对任给 $\delta>0$, 仿 (6) 知存在 $M>0$ 及正整数 N 使当 $n>N$ 时有

$$\begin{aligned}
&P\{|X_nY-XY|\geqslant\varepsilon\}\\
&=P\{|X_nY-XY|\geqslant\varepsilon,\ |Y|=0\}\\
&\quad+P\{|X_nY-XY|\geqslant\varepsilon,\ 0<|Y|<M\}\\
&\quad+P\{|X_nY-XY|\geqslant\varepsilon,\ |Y|\geqslant M\}\\
&\leqslant P\left\{|X_n-X|\geqslant\frac{\varepsilon}{M}\right\}+P\{|Y|\geqslant M\}<\delta
\end{aligned}$$

故知 $X_nY\xrightarrow{P}XY$.

(11) 因 $X_nY_n-XY=X_nY_n-X_nY+X_nY-XY$, 由 (10) 知

$$X_nY\xrightarrow{P}XY$$

用 (6) 及 (10) 中的证明技巧易证

$$X_nY_n-X_nY\xrightarrow{P}0$$

故得 $X_nY_n\xrightarrow{P}XY$.

【评注】本题的主要结论是: 若二个随机变量序列都依概率收敛, 则其和、差、积、商也必依概率收敛, 这结果有时也称斯卢茨基引理.

这里介绍的序列命题及证法体现出把难题分解成较容易的子命题, 再一一解决的典型思路.

27. 设 $X_n\xrightarrow{P}X$, 而 g 是 $\mathbf{R}^1$ 上的连续函数, 试证 $g(X_n)\xrightarrow{P}g(X)$.

证 对任给 $\delta > 0$, 存在 $M > 0$ 及正整数 N_1, 使当 $n > N_1$ 时有

$$P\left\{|X| \geqslant \frac{M}{2}\right\} < \frac{\delta}{4}$$

$$P\left\{|X_n - X| \geqslant \frac{M}{2}\right\} < \frac{\delta}{4}$$

这时

$$\begin{aligned}
&P\{|X_n| \geqslant M\} \\
&= P\left\{|X_n| \geqslant M,\ |X| < \frac{M}{2}\right\} + P\left\{|X_n| \geqslant M,\ |X| \geqslant \frac{M}{2}\right\} \\
&\leqslant P\left\{|X_n - X| \geqslant \frac{M}{2}\right\} + P\left\{|X| \geqslant \frac{M}{2}\right\} \\
&< \frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2}
\end{aligned}$$

又 $g(x)$ 是 $\mathbf{R}^1$ 上的连续函数, 所以 $g(x)$ 在 $[-M,\ M]$ 一致连续, 故对任给 $\varepsilon > 0$, 存在 $\eta > 0$, 使任意 $x_1,\ x_2 \in [-M,\ M]$, 且 $|x_1 - x_2| < \eta$ 时有 $|g(x_1) - g(x_2)| < \varepsilon$. 由 $X_n \xrightarrow{P} X$, 又可知存在正整数 N_2, 使当 $n > N_2$时有

$$P\{|X_n - X| \geqslant \eta\} < \frac{\delta}{4}$$

于是当 $n > N = \max(N_1,\ N_2)$ 时, 就有

$$\begin{aligned}
&P\{|g(X_n) - g(X)| \geqslant \varepsilon\} \\
&\quad \leqslant P\{|g(X_n) - g(X)| \geqslant \varepsilon,\ |X_n| < M,\ |X| < M\} \\
&\qquad + P\{(|X_n| \geqslant M) \cup (|X| \geqslant M)\} \\
&\quad \leqslant P\{|X_n - X| \geqslant \eta\} + P\{|X_n| \geqslant M\} + P\left\{|X| \geqslant \frac{M}{2}\right\} \\
&\quad \leqslant \frac{\delta}{4} + \frac{\delta}{2} + \frac{\delta}{4} = \delta
\end{aligned}$$

所以证得 $g(X_n) \xrightarrow{P} g(X)$.

【评注】这条随机变量函数序列的收敛性结论, 形式普遍, 用途极广.

28. 若 $\{X_n\}$ 是单调下降的正随机变量序列, 且 $X_n \xrightarrow{P} 0$, 试证 $X_n \xrightarrow{a.s.} 0$.

证 由 X_n 为正及单调性知, 对任意 $\varepsilon > 0$,

$$\{|X_k| < \varepsilon\} \subset \{|X_n| < \varepsilon\}, \quad n = k, k+1, \cdots$$

于是

$$\left\{ \bigcap_{n=k}^{\infty} (|X_n| < \varepsilon) \right\} \supset \{|X_k| < \varepsilon\}$$

故

$$P\left\{ \bigcap_{n=k}^{\infty} (|X_n| < \varepsilon) \right\} \geqslant P\{|X_k| < \varepsilon\}$$

从而有

$$1 \geqslant \lim_{k\to\infty} P\left\{ \bigcap_{n=k}^{\infty} (|X_n| < \varepsilon) \right\} \geqslant \lim_{k\to\infty} P\{|X_k| < \varepsilon\} = 1$$

由《概率论基础》(5.4.17) 知 $X_n \xrightarrow{a.s.} 0$.

【评注】单调性假定下, 依概率收敛与以概率 1 收敛等价.

29. 若 $X_1, X_2, \cdots$ 是独立随机变量序列, 其特征函数均为 $\varphi(t)$, μ 是整值随机变量, $P\{\mu = k\} = p_k$, 且与 $\{X_i\}$ 独立, 求 $\eta = X_1 + X_2 + \cdots + X_\mu$ 的特征函数.

解

$$\begin{aligned} f_\eta(t) &= Ee^{\mathrm{i}t\eta} = Ee^{\mathrm{i}t(X_1+\cdots+X_\mu)} \\ &= \sum_{k=1}^{\infty} P\{\mu = k\} E\left[e^{\mathrm{i}t(X_1+\cdots+X_\mu)} \mid \mu = k\right] \\ &= \sum_{k=1}^{\infty} P\{\mu = k\} E\left[e^{\mathrm{i}t(X_1+\cdots+X_k)} \mid \mu = k\right] \\ &= \sum_{k=1}^{\infty} p_k E\left[e^{\mathrm{i}t(X_1+\cdots+X_k)}\right] \\ &= \sum_{k=1}^{\infty} p_k \varphi^k(t) \end{aligned}$$

【评注】请与母函数的相应公式对照, 可见特征函数也能用来处理随机个独立随机变量和的问题, 而且此处 X_i 可以是取任意实数值的随机变量, 因而处理的范围更广.

30. 若 $f(t)$ 是非负定函数, 试证: (1) $f(0)$ 是实的, 且 $f(0) \geqslant 0$; (2) $f(-t) = \overline{f(t)}$; (3) $|f(t)| \leqslant f(0)$.

证 因 $f(t)$ 是非负定函数, 所以对任何实数 $t_1, t_2, \cdots, t_n$, 复数 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 有

$$\sum_{j=1}^{n}\sum_{k=1}^{n} f(t_j - t_k)\lambda_j\overline{\lambda}_k \geqslant 0 \qquad (*)$$

(1) 在 (*) 中令 $n = 1$, $t_1 = 0$, $\lambda_1 = 1$ 得 $f(0) \geqslant 0$, 当然 $f(0)$ 是实的.

(2) 在 (*) 中令 $n = 2$, $t_1 = 0$, $t_2 = t$ 得

$$f(0)|\lambda_1|^2 + f(0)|\lambda_2|^2 + f(-t)\lambda_1\overline{\lambda}_2 + f(t)\overline{\lambda}_1\lambda_2 \geqslant 0 \qquad (**)$$

从而得知

$$f(-t)\lambda_1\overline{\lambda}_2 + f(t)\overline{\lambda}_1\lambda_2 = \overline{f(-t)\lambda_1\overline{\lambda}_2 + f(t)\overline{\lambda}_1\lambda_2}$$

即

$$\left[f(t) - \overline{f(-t)}\right]\overline{\lambda}_1\lambda_2 + \left[f(-t) - \overline{f(t)}\right]\lambda_1\overline{\lambda}_2 = 0$$

由于 λ_1, λ_2 是任意的, 故

$$f(t) = \overline{f(-t)}$$

$$f(-t) = \overline{f(t)}$$

(3) 在 (**) 中令 $\lambda_1 = f(t)$, $\lambda_2 = -|f(t)|$, 利用 $f(-t) = \overline{f(t)}$ 即得

$$2f(0)|f(t)|^2 - |f(t)|^2|f(t)| - |f(t)|^2|f(t)| \geqslant 0$$

若 $|f(t)| \neq 0$, 则由上式得 $f(0) \geqslant |f(t)|$;

若 $|f(t)| = 0$, 由 (1) 知 $f(0) \geqslant |f(t)|$.

【评注】特征函数的性质, 除连续性和规范性 $f(0) = 1$ 外, 都来自非负定性.

31. 某理发店为每个顾客的服务时间服从均值为 1/3 (小时) 的指数分布, 可认为对每个顾客的服务是相互独立的.

(1) 求为对 100 个顾客服务, 总共需要 31 小时至 35 小时的概率;

(2) 以 95% 的概率在 32 小时之内可服务完几个顾客?

(3) 找 Δ, 使该店对 100 个顾客的服务时间在 $(33.33-\Delta, 33.33+\Delta)$ 之间的概率大于 95%.

解 设为顾客的服务时间 $X_i \sim \mathrm{Exp}(3), i=1,2,\cdots,100$, 且相互独立, 按林德伯格–莱维中心极限定理, $X_1+X_2+\cdots+X_{100}$ 近似服从 $N\left(\frac{100}{3}, \frac{100}{9}\right)$.

(1) 求概率

$$
\begin{aligned}
&P\{31 \leqslant X_1+\cdots+X_{100} \leqslant 35\} \\
&=P\left\{\frac{31-\frac{100}{3}}{\frac{10}{3}} \leqslant \frac{X_1+\cdots+X_{100}-\frac{100}{3}}{\frac{10}{3}} \leqslant \frac{35-\frac{100}{3}}{\frac{10}{3}}\right\} \\
&=\Phi(0.5)-\Phi(-0.7) \\
&=0.449\,5
\end{aligned}
$$

(2) **提示:** 求 n.

答 81 个.

(3) **提示:** 求偏差.

答 $\Delta=6.54$ (小时).

【评注】为林德伯格–莱维中心极限定理特设的数值例子, 务必自己完成. 请与频率估计概率的三类计算 (亦可参看习题五 17, 15, 18 题) 问题作比较.

32. 若总体 ξ 的数学期望 $E\xi=m$, $D\xi=\sigma^2$, 抽容量为 n 的样本, 求其平均值 $\bar{\xi}$, 为使 $P\{|\bar{\xi}-m|<0.1\sigma\} \geqslant 95\%$, 问 n 应取多大值?

提示: 大样本场合, 用林德伯格–莱维中心极限定理, 化为正态逼近求 n.

答 n 取 385.

【评注】抽样调查方案设计时的基本计算.

33. 用特征函数法直接证明棣莫弗–拉普拉斯积分极限定理.

证 以 μ_n 记 n 次伯努利试验中成功出现的次数, 则它服从二项分布 $B(n,p)$, 其特征函数 $f_n(t)=\left(q+p\,\mathrm{e}^{\mathrm{i}t}\right)^n$, 由此 $\dfrac{\mu_n-np}{\sqrt{npq}}$ 的特征函数为

$$
\begin{aligned}
g_n(t)&=\left[q+p\,\mathrm{e}^{\frac{\mathrm{i}t}{\sqrt{npq}}}\right]^n\mathrm{e}^{-\frac{np\mathrm{i}t}{\sqrt{npq}}}\\
&=\left[q\,\mathrm{e}^{-\frac{p\mathrm{i}t}{\sqrt{npq}}}+p\,\mathrm{e}^{\frac{q\mathrm{i}t}{\sqrt{npq}}}\right]^n\\
&=\left[q\left(1-\frac{p\mathrm{i}t}{\sqrt{npq}}-\frac{pt^2}{2nq}\right)+p\left(1+\frac{q\mathrm{i}t}{\sqrt{npq}}-\frac{qt^2}{2np}\right)+o\left(\frac{t^2}{n}\right)\right]^n\\
&=\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n
\end{aligned}
$$

可见 $g_n(t)\longrightarrow \mathrm{e}^{-\frac{1}{2}t^2}$, $(n\to\infty)$.

用连续性定理即证得棣莫弗–拉普拉斯极限定理.

【评注】经典的定理, 标准的证明.

34. 若 $\{\xi_n,\ n=1,2,\cdots\}$ 为相互独立随机变量序列, 具有相同分布

$$
P\{\xi_n=1\}=\frac{1}{2},\quad P\{\xi_n=0\}=\frac{1}{2}
$$

而 $\eta_n=\sum\limits_{k=1}^{n}\dfrac{\xi_k}{2^k}$, 试证 η_n 的分布收敛于 $[0,1]$ 上的均匀分布.

证 $\xi_j\ (j=1,2,\cdots)$ 的特征函数为

$$
f_{\xi_j}(t)=\frac{1}{2}\left(\mathrm{e}^{\mathrm{i}t}+1\right)=\frac{1}{2}\left(\mathrm{e}^{\frac{1}{2}\mathrm{i}t}+\mathrm{e}^{-\frac{1}{2}\mathrm{i}t}\right)\mathrm{e}^{\frac{1}{2}\mathrm{i}t}=\cos\frac{t}{2}\,\mathrm{e}^{\frac{1}{2}\mathrm{i}t}
$$

这样 $\frac{\xi_k}{2^k}$ 的特征函数为

$$f_{\frac{\xi_k}{2^k}}(t)=\cos\frac{t}{2^{k+1}}\,\mathrm{e}^{\frac{\mathrm{i}\,t}{2^{k+1}}}$$

从而 η_n 的特征函数为

$$\begin{aligned}f_{\eta_n}(t)&=\cos\frac{t}{2^2}\cos\frac{t}{2^3}\cdots\cos\frac{t}{2^{n+1}}\,\mathrm{e}^{\mathrm{i}\,t\left(\frac{1}{2^2}+\frac{1}{2^3}+\cdots+\frac{1}{2^{n+1}}\right)}\\&=\frac{1}{2^n}\frac{\sin\frac{t}{2}}{\sin\frac{t}{2^{n+1}}}\,\mathrm{e}^{\frac{\mathrm{i}\,t}{2}\left(1-\frac{1}{2^n}\right)}\end{aligned}$$

由此得

$$\lim_{n\to\infty}f_{\eta_n}(t)=\frac{2}{t}\sin\frac{t}{2}\,\mathrm{e}^{\frac{1}{2}\mathrm{i}\,t}=\frac{2}{t}\frac{1}{2\mathrm{i}}\left(\mathrm{e}^{\frac{1}{2}\mathrm{i}\,t}-\mathrm{e}^{-\frac{1}{2}\mathrm{i}\,t}\right)\mathrm{e}^{\frac{1}{2}\mathrm{i}\,t}=\frac{1}{\mathrm{i}\,t}\left(\mathrm{e}^{\mathrm{i}\,t}-1\right)$$

而 $\frac{1}{\mathrm{i}t}\left(\mathrm{e}^{\mathrm{i}\,t}-1\right)$ 是 $[0,1]$ 上均匀分布的特征函数, 故再用连续性定理即证得本题结论.

【评注】本题建立了 $[0,1]$ 上均匀分布与 $[0,1]$ 上的数的二进制展开以及可列重伯努利试验之间的关系. 这个模型导致许多深刻的研究, 包括博雷尔强大数定律. 参看教学札记之二十七 "$[0,1]$ 中数的 g 进制展开与概率论".

35. 用特征函数法证明二项分布的泊松逼近定理.

证 成功概率为 p_n 的二项分布的特征函数为

$$f_n(t)=\left(p_n\,\mathrm{e}^{\mathrm{i}\,t}+q_n\right)^n=\left[1+\frac{np_n(\mathrm{e}^{\mathrm{i}\,t}-1)}{n}\right]^n$$

若当 $n\to\infty$ 时, $np_n\to\lambda$, 则

$$\lim_{n\to\infty}f_n(t)=\exp\left[\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)\right]=f(t)$$

而 $f(t)$ 是泊松分布的特征函数, 故而得证.

【评注】试与课文中用母函数法的证明作对比.

36. 用特征函数法证明, 泊松分布当 $\lambda\to\infty$ 时, 渐近正态分布.

证 设 ξ_λ 服从参数为 λ 的泊松分布, 则其特征函数是

$$f_{\xi_\lambda}(t) = \exp\left[\lambda(\mathrm{e}^{\mathrm{i}\,t} - 1)\right]$$

令 $\eta_\lambda = \dfrac{\xi_\lambda - \lambda}{\sqrt{\lambda}}$, 则 η_λ 的特征函数为

$$\begin{aligned}
f_{\eta_\lambda}(t) &= \mathrm{e}^{-\mathrm{i}\sqrt{\lambda}\,t} f_{\xi_\lambda}\Big(\frac{t}{\sqrt{\lambda}}\Big) \\
&= \exp\Big[-\mathrm{i}\sqrt{\lambda}\,t + \lambda\Big(\mathrm{e}^{\frac{\mathrm{i}\,t}{\sqrt{\lambda}}} - 1\Big)\Big] \\
&= \exp\Big[-\mathrm{i}\sqrt{\lambda}\,t + \mathrm{i}\frac{t}{\sqrt{\lambda}}\lambda - \frac{t^2}{2\lambda}\lambda + O\Big(\frac{1}{\sqrt{\lambda}}\Big)\Big] \\
&= \exp\Big[-\frac{t^2}{2} + O\Big(\frac{1}{\sqrt{\lambda}}\Big)\Big]
\end{aligned}$$

$$\lim_{\lambda\to\infty} f_{\eta_\lambda}(t) = \mathrm{e}^{-\frac{t^2}{2}}$$

所以得泊松分布在 $\lambda \to \infty$ 时渐近正态分布 $N(\lambda, \lambda)$.

【评注】有名的结果, 但也在意料之中.

*37. 若 $\{X_i\}$ 是独立同分布随机变量序列, 其分布分别为: (1) $[-a, a]$ 上的均匀分布; (2) 泊松分布; (3) Γ 分布, 记

$$Y_n = \frac{\sum_{j=1}^{n}(X_j - EX_j)}{\sqrt{\sum_{j=1}^{n} DX_j}}.$$

试计算 Y_n 的特征函数, 并求 $n \to \infty$ 时的极限.

解 (1) 设 X_j 服从 $[-a, a]$ 上的均匀分布, 其特征函数为

$$f_{X_j}(t) = \frac{\sin a\,t}{a\,t}$$

又

$$EX_j = 0, \qquad DX_j = \frac{1}{3}\,a^2, \qquad Y_n = \sum_{j=1}^{n}\frac{\sqrt{3}X_j}{\sqrt{n}a}$$

所以

$$f_{Y_n}(t)=\left[f_{X_j}\left(\sqrt{\frac{3}{n}}\,\frac{t}{a}\right)\right]^n=\left(\frac{\sin\sqrt{\frac{3}{n}}\,t}{\sqrt{\frac{3}{n}}\,t}\right)^n$$

$$=\left[1-\frac{1}{6}\,\frac{3}{n}\,t^2+o\Big(\frac{1}{n}\Big)\right]^n$$

$$=\left[1-\frac{t^2}{2n}+o\Big(\frac{1}{n}\Big)\right]^n\longrightarrow \mathrm{e}^{-\frac{t^2}{2}}\quad(n\to\infty)$$

(2) 设 X_j 服从参数为 λ 的泊松分布，其特征函数为

$$f_{X_j}(t)=\exp\left[\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)\right]$$

又

$$EX_j=\lambda\,,\qquad DX_j=\lambda\,,\qquad Y_n=\sum_{j=1}^{n}\frac{X_j-\lambda}{\sqrt{n\lambda}}$$

所以

$$f_{Y_n}(t)=\left\{\exp\Big(-\mathrm{i}\frac{\sqrt{\lambda}}{\sqrt{n}}\,t\Big)\,\exp\left[\lambda\Big(\mathrm{e}^{\frac{\mathrm{i}\,t}{\sqrt{n\lambda}}}-1\Big)\right]\right\}^n$$

$$=\left\{\exp\Big(-\mathrm{i}\frac{\sqrt{\lambda}}{\sqrt{n}}\,t\Big)\exp\left[\lambda\Big(\frac{\mathrm{i}\,t}{\sqrt{n\lambda}}-\frac{1}{2}\,\frac{t^2}{n\lambda}+o\Big(\frac{1}{n}\Big)\Big)\right]\right\}^n$$

$$=\exp\left[-\mathrm{i}\,t\sqrt{n\lambda}+\mathrm{i}\,t\sqrt{n\lambda}-\frac{t^2}{2}+o(1)\right]$$

$$=\exp\left[-\frac{t^2}{2}+o(1)\right]\longrightarrow \mathrm{e}^{-\frac{t^2}{2}}\quad(n\to\infty)$$

(3) 设 X_j 服从 Γ 分布 $\Gamma(\alpha,\beta)$，其特征函数为

$$f_{X_j}(t)=\Big(1-\frac{\mathrm{i}\,t}{\beta}\Big)^{-\alpha},\quad \beta>0$$

又

$$EX_j=\frac{\alpha}{\beta}\,,\qquad DX_j=\frac{\alpha}{\beta^2}\,,\qquad Y_n=\sum_{j=1}^{n}\frac{X_j-\frac{\alpha}{\beta}}{\sqrt{\frac{n\alpha}{\beta^2}}}$$

所以

$$f_{Y_n}(t)=\left[\exp\left(-\mathrm{i}\sqrt{\frac{\alpha}{n}}\,t\right)\left(1-\mathrm{i}\,\frac{t}{\sqrt{n\alpha}}\right)^{-\alpha}\right]^n$$

$$=\exp\left(-\mathrm{i}\,t\sqrt{n\alpha}\right)\left(1-\mathrm{i}\,\frac{t}{\sqrt{n\alpha}}\right)^{-n\alpha}$$

$$\ln f_{Y_n}(t)=-\mathrm{i}\,t\sqrt{n\alpha}-n\alpha\ln\left(1-\mathrm{i}\,\frac{t}{\sqrt{n\alpha}}\right)$$

$$=-\mathrm{i}\,t\sqrt{n\alpha}-n\alpha\left[-\frac{\mathrm{i}\,t}{\sqrt{n\alpha}}+\frac{t^2}{2n\alpha}+o\left(\frac{1}{n}\right)\right]$$

$$\longrightarrow -\frac{t^2}{2}\quad (n\to\infty)$$

从而亦有 $f_{Y_n}(t)\longrightarrow \mathrm{e}^{-\frac{t^2}{2}}\quad (n\to\infty)$.

【评注】对不同分布, 用标准方法, 达相同结论. 充分显示出特征函数法在证明中心极限定理中的威力. 第五章 §3 例 4 为 (1) 提供直观; 注意到泊松分布的再生性可知 (2) 与上题关系密切.

38. 设 $\{X_n\}$ 独立同分布, $P\{X_n=2^{k-2\ln k}\}=2^{-k}$ $(k=1,2,\cdots)$, 则大数定律成立.

证
$$EX_i=\sum_{k=1}^{\infty}2^{k-2\ln k}2^{-k}=\sum_{k=1}^{\infty}4^{-\ln k}=\sum_{k=1}^{\infty}k^{-\ln 4}$$

因 $\ln 4>1$, 所以 $\sum\limits_{k=1}^{\infty}k^{-\ln 4}$ 收敛, 即 EX_i 存在, 由辛钦大数定律即知对 $\{X_n\}$ 大数定律成立.

【评注】独立同分布场合首先想到辛钦大数定律.

*39. 若 $\{X_i\}$ 是相互独立的随机变量序列, 均服从 $N(0,1)$, 试证

$$W_n=\sqrt{n}\frac{X_1+\cdots+X_n}{X_1^2+\cdots+X_n^2}\quad 及\quad U_n=\frac{X_1+\cdots+X_n}{\sqrt{X_1^2+\cdots+X_n^2}}$$

渐近正态分布 $N(0,1)$.

证 (1) 设 $\xi_n=\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}X_i$, $\eta_n=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^2$, 由题设条件用正

态分布的再生性得

$$\xi_n \sim N(0,1)$$

用辛钦大数定律得

$$\eta_n \xrightarrow{P} 1$$

又

$$W_n = \frac{\dfrac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i}{\dfrac{1}{n}\sum_{i=1}^{n} X_i^2} = \frac{\xi_n}{\eta_n}$$

由斯卢茨基定理 (见本章习题 25) 知 W_n 渐近正态分布.

(2) 由 (1) 得 $\eta_n \xrightarrow{P} 1$, 又由本章习题 27 可知 $\sqrt{\eta_n} \xrightarrow{P} 1$ 且

$$U_n = \frac{\dfrac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i}{\sqrt{\dfrac{1}{n}\sum_{i=1}^{n} X_i^2}} = \frac{\xi_n}{\sqrt{\eta_n}}$$

同样用斯卢茨基定理得 U_n 渐近正态分布.

【评注】同类题型问题甚多, 在数理统计中有用, 证明方法也很有教益.

*40. 设 $X_1, X_2, \cdots$ 是独立随机变量序列, 均服从 $[0, 1]$ 均匀分布, 令

$$Z_n = \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}}$$

试证 $Z_n \xrightarrow{P} C$, 这里 C 是常数, 并求 C.

证

$$\ln Z_n = \frac{1}{n}\sum_{i=1}^{n} \ln X_i$$

又

$$E\ln X_i = \int_0^1 \ln x \,\mathrm{d}x = \left(x\ln x - x\right)\Big|_0^1 = -1$$

由辛钦大数定律知 $\ln Z_n \xrightarrow{P} -1$, 因而利用本章习题 27 即得 $Z_n \xrightarrow{P} \mathrm{e}^{-1}\quad (n\to\infty)$. 易见 $C=\mathrm{e}^{-1}$.

【评注】独立 $[0,1]$ 均匀分布随机变量的几何平均值必收敛于 e^{-1}.

*41. 若 $\{X_i\}$ 是独立同分布随机变量序列, $EX_i=m$, 而 $f(x)$ 是一个有界的连续函数, 试证

$$\lim_{n\to\infty} E\left[f\left(\frac{X_1+\cdots+X_n}{n}\right)\right]=f(m)$$

证 由科尔莫戈罗夫强大数定律得

$$\frac{1}{n}(X_1+X_2+\cdots+X_n) \xrightarrow{a.s.} EX_i=m \quad (n\to\infty)$$

又因 $f(x)$ 有界, 由控制收敛定理和 $f(x)$ 的连续性, 则有

$$\begin{aligned}\lim_{n\to\infty} E\left[f\left(\frac{X_1+\cdots+X_n}{n}\right)\right] &= E\left[\lim_{n\to\infty} f\left(\frac{X_1+\cdots+X_n}{n}\right)\right]\\ &= E\left[f\left(\lim_{n\to\infty}\frac{X_1+\cdots+X_n}{n}\right)\right]\\ &= f(m)\end{aligned}$$

【评注】请特别注意所给条件的使用.

*42. 若 $\{X_i\}$ 是独立同分布, 具有有限二阶矩的随机变量序列, 试证

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{P} EX_1$$

证 记 $EX_i=a$, $DX_i=\sigma^2$, 则

$$E\left(\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i\right)=\frac{2}{n(n+1)}\sum_{i=1}^{n} iEX_i=a$$

$$\begin{aligned}E\left(\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i-a\right)^2 &= \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n} i^2\sigma^2\\ &= \frac{2(2n+1)\sigma^2}{3n(n+1)} \longrightarrow 0 \quad (n\to\infty)\end{aligned}$$

故

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{\text{均方收敛}} a$$

因此由《概率论基础》定理 5.2.8 推出

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i \xrightarrow{P} a = EX_1$$

【评注】通过 r 阶收敛证明依概率收敛的一个例子, 常用于数理统计中相合性的证明, 本题亦有统计学背景.

*43. 设 $X_1, X_2, \cdots$ 相互独立, 均服从柯西分布 $p(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$, 试证它们不满足格涅坚科关于大数定律的充要条件 (见本章习题 9), 即要指出, 当 $n \to \infty$ 时

$$E\left[\frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n^2+\left(\sum_{i=1}^{n} X_i\right)^2}\right] \not\to 0$$

证 由柯西分布的再生性(参看第四章习题 *49) 知 $\sum_{i=1}^{n} X_i$ 的密度函数为 $\frac{1}{\pi} \cdot \frac{n}{n^2+x^2}$, 因而

$$\begin{aligned}
E\left[\frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n^2+\left(\sum_{i=1}^{n} X_i\right)^2}\right] &= \int_{-\infty}^{\infty} \frac{x^2}{n^2+x^2} \cdot \frac{1}{\pi} \cdot \frac{n}{n^2+x^2}\, \mathrm{d}x \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{u^2}{(1+u^2)^2}\, \mathrm{d}u \quad (x = nu) \\
&= -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{u}{2}\, \mathrm{d}\frac{1}{1+u^2}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2\pi}\,\frac{u}{1+u^2}\bigg|_{-\infty}^{\infty} + \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1}{1+u^2}\,\mathrm{d}u \\
&= \frac{1}{2\pi}\arctan u\bigg|_{-\infty}^{\infty} \\
&= \frac{1}{2\pi}\Big(\frac{\pi}{2}+\frac{\pi}{2}\Big) \\
&= \frac{1}{2} \not\to 0 \quad (n\to\infty)
\end{aligned}
$$

【评注】柯西分布奇怪性质之一是其平均值 $\dfrac{X_1+\cdots+X_n}{n}$ 与 X_1 同分布 (见第四章习题 51), 因此不可能服从大数定律. 另外它根本不存在数学期望, 大数定律这时按《概率论基础》(5.1.9) 定义.

*44. (魏尔斯特拉斯定理的概率论证明) 设 $f(x)$ 是 $[0,1]$ 上连续函数, 利用概率论方法证明: 必存在多项式序列 $\{B_n(x)\}$, 在 $[0,1]$ 上一致收敛于 $f(x)$. (提示: 定义伯恩斯坦多项式

$$
B_n(x) = \sum_{m=0}^{n}\binom{n}{m}x^m(1-x)^{n-m}f\Big(\frac{m}{n}\Big)
$$

并利用大数定律.)

证 $f(x)$ 在 $[0,1]$ 上有界且一致连续, 因而存在 $M>0$, 使对一切 $0\leqslant x\leqslant 1$ 有 $|f(x)|\leqslant M$, 并且对任意 $\varepsilon>0$, 存在 $\delta>0$, 对任意 $0\leqslant x,\ y\leqslant 1$, 只要 $|x-y|<\delta$, 就有 $|f(x)-f(y)|<\dfrac{\varepsilon}{2}$.

依提示定义 $B_n(x)$, 容易验证, $B_n(0)=f(0)$, $B_n(1)=f(1)$, 因而只要考虑 $0<x<1$. 对每一固定 x, 设想一伯努利实验, 使事件 A 在每次试验中出现的概率恰为 x, 并以 η_n 表示前 n 次试验中 A 出现的次数, 则 $B_n(x)=Ef\Big(\dfrac{\eta_n}{n}\Big)$.

对上述的 δ, 由伯努利大数定律有

$$
\lim_{n\to\infty}P\Big\{\Big|\frac{\eta_n}{n}-x\Big|\geqslant\delta\Big\}=0
$$

又因

$$
\begin{aligned}
f(x)-B_n(x) &= \sum_{m=0}^{n}\binom{n}{m}x^m(1-x)^{n-m}\Big[f(x)-f\Big(\frac{m}{n}\Big)\Big] \\
&= E\Big[f(x)-f\Big(\frac{\eta_n}{n}\Big)\Big] \\
&= P\Big\{\Big|\frac{\eta_n}{n}-x\Big|<\delta\Big\}E\Big[f(x)-f\Big(\frac{\eta_n}{n}\Big)\Big|\Big|\frac{\eta_n}{n}-x\Big|<\delta\Big] \\
&\quad + P\Big\{\Big|\frac{\eta_n}{n}-x\Big|\geqslant\delta\Big\}E\Big[f(x)-f\Big(\frac{\eta_n}{n}\Big)\Big|\Big|\frac{\eta_n}{n}-x\Big|\geqslant\delta\Big]
\end{aligned}
$$

于是

$$
|f(x)-B_n(x)|\leqslant 1\cdot\frac{\varepsilon}{2}+P\Big\{\Big|\frac{\eta_n}{n}-x\Big|\geqslant\delta\Big\}\cdot 2M \qquad (*)
$$

这就得到 $\lim\limits_{n\to\infty}B_n(x)=f(x)$ 对每个 $0\leqslant x\leqslant 1$ 成立.

利用切比雪夫不等式

$$
P\Big\{\Big|\frac{\eta_n}{n}-x\Big|\geqslant\delta\Big\}\leqslant\frac{D\eta_n}{n^2\delta^2}=\frac{nx(1-x)}{n^2\delta^2}\leqslant\frac{1}{4n\delta^2}
$$

当 $n>\dfrac{M}{\varepsilon\delta^2}$时, 由上式及 $(*)$ 式得

$$
\sup_{0\leqslant x\leqslant 1}|f(x)-B_n(x)|<\varepsilon
$$

这就证得了 $B_n(x)$ 在 $[0,1]$ 上一致收敛于 $f(x)$.

【评注】著名的分析定理, 巧妙的概率证明.

45. 设 $\{X_n\}$ 是独立随机变量序列, 试证 $X_n\xrightarrow{a.s.}0$ 的充要条件为对任意 $\varepsilon>0$, 有 $\sum\limits_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}<\infty$.

证 充分性: 因为 $\sum\limits_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}<\infty$, 则由博雷尔 - 康特立引理 (i) 知 $P\Big\{\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=k}^{\infty}(|X_n|\geqslant\varepsilon)\Big\}=0$, 这正是表明 $X_n\xrightarrow{a.s.}0\ (n\to\infty)$.

必要性: 若 $X_n \xrightarrow{a.s.} 0$, 则应有 $P\Big\{\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=k}^{\infty}(|X_n|\geqslant\varepsilon)\Big\}=0$.

用反证法, 假设 $\sum\limits_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}=\infty$, 则由博雷尔 - 康特立引理 (ii) 及 $\{X_n\}$ 的独立性得 $P\Big\{\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=k}^{\infty}(|X_n|\geqslant\varepsilon)\Big\}=1$, 这就导致了矛盾. 所以独立随机变量序列 $X_n \xrightarrow{a.s.} 0$ 时, 有 $\sum\limits_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}<\infty$.

【评注】给出独立随机变量序列以概率 1 收敛的一个简明而实用的充要条件. 证明过程也显示博雷尔–康特立引理的强大威力.

46. 试证独立同分布随机变量序列, 若存在有限的四阶中心矩, 则强大数定律成立.

证 因存在四阶中心矩, 故数学期望和方差亦存在, 由独立同分布场合的科尔莫戈罗夫强大数定律即得结论.

也可用博雷尔–康特立引理来证. 设 $EX_i=a$, $DX_i=\sigma^2$, $E(X_i-a)^4=c_4$, 由马尔可夫不等式得

$$\begin{aligned}P\Big\{\Big|\frac{1}{n}\sum_{i=1}^{n}X_i-a\Big|\geqslant\varepsilon\Big\}&\leqslant\frac{1}{\varepsilon^4}E\Big[\frac{1}{n}\sum_{i=1}^{n}(X_i-a)\Big]^4\\&=\frac{1}{n^4\varepsilon^4}\Big\{E\Big[\sum_{i=1}^{n}(X_i-a)^4\Big]+\binom{4}{2}E\Big[\sum_{i\neq j}(X_i-a)^2(X_j-a)^2\Big]\Big\}\\&=\frac{1}{n^4\varepsilon^4}\Big[nc_4+6\cdot\frac{n(n-1)}{2}\sigma^4\Big]\\&\leqslant\frac{1+3\sigma^4}{n^2\varepsilon^4}\end{aligned}$$

其中因独立性有 $E(X_i-a)(X_j-a)^3=0\ (i\neq j)$, 最后一步成立是因为 c_4 有限, 当 n 充分大时定有 $c_4<n$, 而 $\sum\limits_{n=1}^{\infty}\frac{1+3\sigma^4}{n^2\varepsilon^4}<\infty$,

所以

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \geqslant \varepsilon\right\} < \infty \quad .$$

由博雷尔–康特立引理

$$P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| \geqslant \varepsilon\right\} = 0$$

所以 $\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{a.s.} a$, 强大数定律成立.

【评注】本题的证明平行于博雷尔强大数定律. 使用博雷尔–康特立引理, 要证明相应的级数收敛, 用切比雪夫不等式或二阶矩是不够的, 这时用了马尔可夫不等式与四阶矩, 顺利完成.

总之, 博雷尔–康特立引理和马尔可夫不等式是证明以概率 1 收敛的主要工具.

顺便指出, 自科尔莫戈罗夫建立独立同分布场合的强大数定律, 数学期望的存在性成为充要条件, 只需简单加以验证即可. 因此本题的命题含义当然是要求从最基本的事实出发证明历史上早就发现而有意义的一个强大数定律, 也即本题是插在《概率论基础》§4 第二小节之后, 第三小节之前的一道练习题.

47. 本章习题 6 的各个独立随机变量序列是否满足强大数定律?

解 (1) 注意到 $P\{X_k = \pm 2^k\} = \frac{1}{2}$, $EX_k = 0$. 假定 $\{X_k\}$ 满足强大数定律, 则有

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{a.s.} 0$$

亦有

$$\frac{1}{n}(X_1 + \cdots + X_{n-1}) = \frac{n-1}{n} \cdot \frac{1}{n-1}(X_1 + \cdots + X_{n-1}) \xrightarrow{a.s.} 0$$

由上述两式得

$$\frac{1}{n}X_n \xrightarrow{a.s.} 0$$

但由 $\{X_k\}$ 的分布, 以概率 1 有 $\left|\frac{X_n}{n}\right| = \frac{2^n}{n} > 1$, 这显然与 $\frac{1}{n}X_n \xrightarrow{a.s.} 0$ 矛盾, 因而对这个随机序列强大数定律不成立.

(2) $DX_k = 1$, 由于

$$\sum_{k=1}^{\infty} \frac{DX_k}{k^2} = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

所以本随机序列满足科尔莫戈罗夫强大数定律成立的条件.

(3) 注意到本章习题 45 的结论, 对独立随机变量序列 $\{X_n\}$,

$$X_n \xrightarrow{a.s.} 0 \iff \forall\, \varepsilon > 0, \quad \sum_{n=1}^{\infty} P\{|X_n| \geqslant \varepsilon\} < \infty$$

取 $\varepsilon_0 = \frac{1}{2}$, 则

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left\{\frac{1}{n}|X_n| \geqslant \frac{1}{2}\right\} &\geqslant \sum_{n=1}^{\infty} P\left\{\frac{1}{n}|X_n| = 1\right\} \\
&= \sum_{n=1}^{\infty} P\{|X_n| = n\} \\
&= \sum_{n=1}^{\infty} n^{-\frac{1}{2}} = \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}
\end{aligned}$$

发散, 所以 $\frac{1}{n}X_n \stackrel{a.s.}{\not\longrightarrow} 0$, 因而推出 $\frac{1}{n}(X_1 + \cdots + X_n) \stackrel{a.s.}{\not\longrightarrow} 0$, 故这个随机序列不满足强大数定律.

【评注】独立而不同分布场合, 强大数定律成立并未找到充要条件, 而科尔莫戈罗夫条件 $\sum_{k=1}^{\infty} \frac{D\xi_k}{k^2} < \infty$ 是目前找到的最好充分条件, 因此要验证成立, 就验证它. 若要验证不成立, 只好各显神通, 如本题 (1) 及 (3).

48. 举例说明博雷尔–康特立引理 (i) 之逆不成立.

解 设 $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$, 且 $P(A_n) = \frac{1}{n}$, 则

$$
\begin{aligned}
P\left\{\overline{\lim_{n\to\infty}} A_n\right\} &= P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right\} \\
&= P\left\{\bigcap_{n=1}^{\infty} A_n\right\} \\
&= \lim_{n\to\infty} P(A_n) \\
&= \lim_{n\to\infty} \frac{1}{n} = 0
\end{aligned}
$$

然而

$$
\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty
$$

故博雷尔–康特立引理 (i) 之逆并不成立.

【评注】一定要举非独立随机事件序列的例子! 为什么?

49. 设 $\{X_n\}$ 是相互独立且具有有限方差的随机变量序列, 若

$$
\sum_{n=1}^{\infty} \frac{DX_n}{n^2} < \infty
$$

则必有

$$
\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} DX_k = 0
$$

证 因 $\sum_{n=1}^{\infty} \frac{DX_n}{n^2} < \infty$, 则对任给 $\varepsilon > 0$, 存在正整数 N, 有

$$
\sum_{k=N+1}^{\infty} \frac{DX_k}{k^2} < \varepsilon
$$

当 $n > N$,

$$
0 \leqslant \frac{1}{n^2} \sum_{k=1}^{n} DX_k
$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{k=1}^{N} DX_k + \frac{1}{n^2}\sum_{k=N+1}^{n} DX_k \\
&\leqslant \frac{1}{n^2}\sum_{k=1}^{N} DX_k + \sum_{k=N+1}^{n} \frac{DX_k}{k^2} \\
&\leqslant \frac{1}{n^2}\sum_{k=1}^{N} DX_k + \sum_{k=N+1}^{\infty} \frac{DX_k}{k^2} \\
&\leqslant \frac{1}{n^2}\sum_{k=1}^{N} DX_k + \varepsilon
\end{aligned}
$$

对固定的 N, 当 n 充分大时, 上式右边第一项可以小于 ε, 故而有

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n} DX_k = 0$$

【评注】这是一条纯分析的结论, 但是在独立和的大数定律中, 前者是科尔莫戈罗夫关于强大数定律成立的条件, 而后者是马尔可夫关于 (弱) 大数定律成立的条件.

*50. 设 $f(x)$ 和 $g(x)$ 在闭区间 $[0, 1]$ 上连续, 且满足 $0 \leqslant f(x) < Cg(x)$, 这里 C 是一个正常数, 则成立

$$
\begin{aligned}
&\lim_{n\to\infty}\int_0^1\int_0^1\cdots\int_0^1 \frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{g(x_1)+g(x_2)+\cdots+g(x_n)}\,\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n \\
&\quad = \frac{\displaystyle\int_0^1 f(x)\mathrm{d}x}{\displaystyle\int_0^1 g(x)\mathrm{d}x}
\end{aligned}
$$

证 设独立随机变量序列 $\{\xi_n\}$ 之每个变量均服从 $[0,1]$ 上均匀分布, 因 $f(x)$, $g(x)$ 在闭区间 $[0,1]$ 上连续, 故积分 $\int_0^1 f(x)\,\mathrm{d}x$, $\int_0^1 g(x)\,\mathrm{d}x$ 存在. 因而由佚名统计学家公式得

$$Ef(\xi_1) = \int_0^1 f(x)\,\mathrm{d}x$$

亦存在有限, 同样 $Eg(\xi_1)$ 也存在有限.

于是独立同分布随机变量序列 $\{f(\xi_n)\}$ 满足强大数定律, 有

$$P\Big\{\lim_{n\to\infty}\frac{1}{n}\big(f(\xi_1)+\cdots+f(\xi_n)\big)=Ef(\xi_1)\Big\}=1$$

同样有

$$P\Big\{\lim_{n\to\infty}\frac{1}{n}\big(g(\xi_1)+\cdots+g(\xi_n)\big)=Eg(\xi_1)\Big\}=1$$

作

$$\eta_n=\frac{f(\xi_1)+\cdots+f(\xi_n)}{g(\xi_1)+\cdots+g(\xi_n)}$$

由 $0\leqslant f(x)\leqslant Cg(x)$ 知 $0\leqslant\eta_n\leqslant C$, 且 ξ_n, $f(\xi_n)$, $g(\xi_n)$, η_n $(n=1,2,\cdots)$ 是定义在同一概率空间 $(\Omega,\mathscr{F},P)$ 上的随机变量, 故而由多维佚名统计学家公式知

$$\begin{aligned}E\eta_n&=E\left[\frac{f(\xi_1)+\cdots+f(\xi_n)}{g(\xi_1)+\cdots+g(\xi_n)}\right]\\&=\int_0^1\cdots\int_0^1\frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}\,\mathrm{d}x_1\cdots\mathrm{d}x_n\end{aligned}$$

于是

$$\begin{aligned}&\lim_{n\to\infty}\int_0^1\cdots\int_0^1\frac{f(x_1)+\cdots+f(x_n)}{g(x_1)+\cdots+g(x_n)}\,\mathrm{d}x_1\cdots\mathrm{d}x_n\\&\quad=\lim_{n\to\infty}E\eta_n=E\big(\lim_{n\to\infty}\eta_n\big)\ (\text{控制收敛定理})\\&\quad=E\left[\lim_{n\to\infty}\frac{\frac{1}{n}\big(f(\xi_1)+\cdots+f(\xi_n)\big)}{\frac{1}{n}\big(g(\xi_1)+\cdots+g(\xi_n)\big)}\right]\\&\quad=\frac{Ef(\xi_1)}{Eg(\xi_1)}=\frac{\int_0^1 f(x)\,\mathrm{d}x}{\int_0^1 g(x)\,\mathrm{d}x}\end{aligned}$$

【评注】一个趣味式的分析问题用概率论的方法给予严格证明.

51. 设 $X_1, X_2, \cdots$ 是独立随机变量序列, 对它成立中心极限定理, 则对 $\{X_n\}$ 成立大数定律的充要条件为 $D(X_1+\cdots+X_n)=o(n^2)$.

证 充分性: 设 $D(X_1+\cdots+X_n)=o(n^2)$, 则由切比雪夫不等式得

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(X_i-EX_i)\right|\geqslant\varepsilon\right\}\leqslant\frac{D(X_1+\cdots+X_n)}{\varepsilon^2 n^2}\longrightarrow 0\quad(n\to\infty)$$

所以大数定律成立.

必要性: 设对 $\{X_n\}$ 成立中心极限定理, 即对任意的 $A>0$, 有

$$\lim_{n\to\infty}P\left\{\frac{1}{B_n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<A\right\}=\frac{1}{\sqrt{2\pi}}\int_{-A}^{A}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t \tag{1}$$

其中 $B_n=\sqrt{D(X_1+\cdots+X_n)}$, 又对 $\{X_n\}$ 成立大数定律, 即对任给 $\varepsilon>0$, 有

$$\lim_{n\to\infty}P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\right\}=1 \tag{2}$$

注意到

$$\begin{aligned}
&P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\right\}\\
&\quad=P\left\{\frac{B_n}{n}\cdot\frac{1}{B_n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\right\}\\
&\quad=P\left\{\frac{1}{B_n}\left|\sum_{i=1}^{n}(X_i-EX_i)\right|<\varepsilon\cdot\frac{n}{B_n}\right\}
\end{aligned}$$

由此利用 (1), (2) 两式可得, 当 $n\to\infty$ 时应有 $\dfrac{\varepsilon n}{B_n}\longrightarrow\infty$, 即 $\dfrac{B_n^2}{n^2}\longrightarrow 0$, 从而得 $D(X_1+\cdots+X_n)=o(n^2)$.

【评注】$D(X_1+\cdots+X_n)=o(n^2)$ 就是著名的马尔可夫条件 (5.1.13), 它是大数定律成立的充分条件. 本题表明, 在独立和场合, 若成立中心极限定理, 则它也是必要条件.

52. 设 $X_1, X_2, \cdots$ 是独立同分布随机变量序列, 且 $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k$ 对每一个 $n=1,2,\cdots$ 有相同分布, 那么, 若 $EX_i=0$, $DX_i=1$, 则 X_i 必须是 $N(0,1)$ 变量.

证 当 $n=1$ 时, $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k$ 即为 X_1, 所以如果 $f(t)$ 是 X_i 的特征函数, 那么根据题设 $f(t)$ 也是 $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k$ 的特征函数, 并且

$$f(t)=E\exp\left\{\mathrm{i}\,\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k\, t\right\}=\left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

对每个 n 均成立, 又

$$f\left(\frac{t}{\sqrt{n}}\right)=1+\mathrm{i}EX_1\frac{t}{\sqrt{n}}+\frac{\mathrm{i}^2EX_1^2}{2}\cdot\frac{t^2}{n}+o\left(\frac{1}{n}\right)=1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)$$

则有

$$\left[f\left(\frac{1}{\sqrt{n}}\right)\right]^n=\left[1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right]^n\longrightarrow \mathrm{e}^{-\frac{t^2}{2}}\quad (n\to\infty)$$

$$f(t)=\lim_{n\to\infty}\left[f\left(\frac{t}{\sqrt{n}}\right)\right]^n=\mathrm{e}^{-\frac{t^2}{2}}$$

所以 $X_i\sim N(0,1)$.

【评注】正态分布的又一个刻画性质.

53. 设 $\{X_k\}$ 是独立随机变量序列, 且 X_k 服从 $N(0,2^{-k})$, 试证序列 $\{X_k\}$: (1) 成立中心极限定理; (2) 不满足费勒条件; (3) 不满足林德伯格条件, 从而说明林德伯格条件并不是中心极限定理成立的必要条件.

证 (1) 因为 X_k 服从正态分布, 所以 $\frac{1}{B_n}\sum_{k=1}^{n} X_k$ 对每个 n 都服从 $N(0,1)$, 因此中心极限定理成立.

(2)
$$B_n^2 = \sum_{k=1}^{n} b_k^2 < \sum_{k=1}^{\infty} b_k^2 = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1$$

$$B_n^2 \not\to \infty \ (n \to \infty)$$

所以不满足费勒条件之《概率论基础》(5.5.4).

(3) $a_k = EX_k = 0$, $B_n \uparrow 1$. 取 $\tau = 1$ 得

$$\sum_{k=1}^{n} \int_{|x|>\tau B_n} x^2 \mathrm{d}F_k(x) > \int_{|x|>1} x^2 \mathrm{d}F_1(x) = C > 0$$

因此
$$\lim_{n\to\infty} \frac{1}{B_n} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2 \mathrm{d}F_k(x) \neq 0$$

故知不满足林德伯格条件.

【评注】本题以反例形式对林德伯格–费勒中心极限定理作出补充说明: (1) 林德伯格条件不是中心极限定理成立的必要条件; (2) 费勒条件不能由中心极限定理推出.

54. 若 $\{X_k\}$ 是独立随机变量序列, X_1 服从 $[-1,1]$ 均匀分布, 对 $k=2,3,\cdots$, X_k 服从 $N(0,2^{k-1})$, 证明对 $\{X_k\}$ 成立中心极限定理但不满足费勒条件.

证 $EX_k = 0$, $B_n^2 = \dfrac{1}{3} + \sum\limits_{k=2}^{n} 2^{k-1} = 2^n - \dfrac{5}{3}$, 又 $\sum\limits_{k=1}^{n} X_k$ 的特征函数为

$$\begin{aligned} f(t) = \prod_{k=1}^{n} f_{X_k}(t) &= \frac{\sin t}{t} \prod_{k=2}^{n} \exp\left\{-\frac{1}{2}\, 2^{k-1} t^2\right\} \\ &= \frac{\sin t}{t} \exp\left\{-\frac{1}{2}(2^n-2)\, t^2\right\} \end{aligned}$$

由此得 $\zeta_n = \dfrac{1}{\sqrt{2^n - 5/3}} \sum_{k=1}^{n} X_k$ 的特征函数为

$$g(t) = \sin \frac{t}{\sqrt{2^n - 5/3}} \cdot \frac{\sqrt{2^n - 5/3}}{t} \exp\left\{ -\frac{(2^n - 2)\, t^2}{2(2^n - 5/3)} \right\}$$
$$\longrightarrow \mathrm{e}^{-\frac{t^2}{2}} \quad (n \to \infty)$$

因而得 $\zeta_n \xrightarrow{L} N(0,1)$, 所以中心极限定理成立.

然而尽管

$$B_n^2 = 2^n - \frac{5}{3} \longrightarrow \infty$$

但

$$\frac{b_n^2}{B_n^2} = \frac{2^{n-1}}{2^n - 5/3} \longrightarrow \frac{1}{2} \neq 0 \quad (n \to \infty)$$

所以不满足费勒条件之《概率论基础》(5.5.5).

【评注】当然本例中的 $\{X_k\}$ 也不满足林德伯格条件, 因此本例要说明的问题与上例完全相同, 差别在于上例因 $B_n^2 \not\to \infty$ 而不满足费勒第一个条件, 而本例则不满足《概率论基础》定理 5.5.1 中的费勒第二个条件.

通过上题与本题, 让读者对林德伯格–费勒中心极限定理的成立条件有更好的理解.

*55. 在泊松试验中, 第 i 次试验时事件 A 出现的概率为 p_i, 不出现的概率为 q_i, 各次试验是独立的, 以 ν_n 记前 n 次试验中事件 A 出现的次数, 试证:

(1) $\dfrac{\nu_n - E\nu_n}{n} \xrightarrow{P} 0$;

(2) 对 $\{\nu_n\}$ 成立中心极限定理的充要条件是 $\sum\limits_{i=1}^{\infty} p_i q_i = +\infty$.

证 记

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次 } A \text{ 出现} \\ 0, & \text{第 } i \text{ 次 } A \text{ 不出现} \end{cases}$$

则

$$P\{X_i=1\}=p_i\ ,\quad P\{X_i=0\}=q_i$$

而 $\nu_n=X_1+\cdots+X_n$.

(1) 易知 $p_iq_i=p_i(1-p_i)\leqslant\dfrac{1}{4}$, 因而 $D\nu_n=\sum\limits_{i=1}^{n}p_iq_i\leqslant\dfrac{1}{4}n$. 由切比雪夫不等式得

$$P\left\{\left|\frac{\nu_n-E\nu_n}{n}\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}\frac{D\nu_n}{n^2}\leqslant\frac{1}{4\varepsilon^2 n}\longrightarrow 0$$

所以

$$\frac{\nu_n-E\nu_n}{n}\xrightarrow{P}0\quad(n\to\infty)$$

这就是教科书中提到的泊松大数定律.

(2) 充分性:

易知 $\max\limits_{1\leqslant j\leqslant n}|X_j|\leqslant 1$, 用假设 $B_n^2=D\nu_n=\sum\limits_{i=1}^{n}p_iq_i\longrightarrow\infty$, 得

$$\lim_{n\to\infty}\frac{1}{B_n}=0$$

由《概率论基础》定理 5.5.3 即得对 ν_n 成立中心极限定理.

必要性: 用反证法. 假设 $\sum\limits_{i=1}^{\infty}p_iq_i$ 收敛, 记 $B^2=\sum\limits_{i=1}^{\infty}p_iq_i$, $B_n^2=\sum\limits_{i=1}^{n}p_iq_i$, 易知 $\{B_n^2\}$ 单调上升, $B^2=\lim\limits_{n\to\infty}B_n^2$. 又记

$$Z_n=\frac{\nu_n-\sum\limits_{i=1}^{n}p_i}{\sqrt{\sum\limits_{i=1}^{n}p_iq_i}}=\frac{\nu_n-E\nu_n}{B_n}$$

ν_n 取 $0,1,2,\cdots,n$, 则 Z_n 可取 $\dfrac{k-E\nu_n}{B_n},\quad k=0,1,2,\cdots,n$, 相邻值之间隔为 $\dfrac{1}{B_n}$. 由 $\{B_n^2\}$ 单调上升并以 B^2 为极限, 可知对 $\forall n$, $B_n\leqslant B$. 于是就有, 对所有的 n, Z_n 所有取值的间隔不小于 $\dfrac{1}{B}$.

这样的 Z_n 的分布函数不可能收敛到标准正态分布. 如若不然, 假定 Z_n 依分布收敛到标准正态分布, 则由本章习题 *20 知这样的收敛是一致的, 于是对每个 n 总可取到区间 $I_n = (a_n, b_n)$, 例如可取

$$a_n = \frac{\left[\sum_{i=1}^{n} p_i\right] - \sum_{i=1}^{n} p_i}{B_n}, \quad b_n = a_n + \frac{1}{2B}$$

这样的 I_n 中不含 Z_n 所有可取的数值, 即

$$P\{a_n < Z_n < b_n\} = 0$$

但由 Z_n 的分布函数一致地收敛于标准正态分布, 推得

$$P\{a_n < Z_n < b_n\} - \frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u \longrightarrow 0 \quad (n \to \infty)$$

由此得

$$\frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u \longrightarrow 0 \quad (n \to \infty)$$

这与任意 n, $b_n - a_n = \dfrac{1}{2B}$ 相矛盾, 所以 Z_n 不可能依分布收敛于标准正态分布, 即对 $\{\nu_n\}$ 不成立中心极限定理.

【评注】对泊松试验完全搞清两种极限定理成立的条件.

56. 设 $\{X_k\}$ 是独立随机变量序列, X_k 服从 $[-k, k]$ 均匀分布, 问对 $\{X_k\}$ 能否用中心极限定理?

解

$$a_k = EX_k = 0, \quad DX_k = \frac{1}{3}k^2$$

$$B_n^2 = \sum_{k=1}^{n} DX_k = \frac{1}{3}\sum_{k=1}^{n} k^2 = \frac{1}{18}n(n+1)(2n+1)$$

对任意 $\tau > 0$, 有

$$\frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x| > \tau B_n} x^2 \mathrm{d}F_k(x)$$

$$\leqslant \frac{1}{B_n^2 \tau B_n} \sum_{k=1}^{n} \int_{|x|>\tau B_n} |x| x^2 f_k(x) \mathrm{d}x$$

$$\leqslant \frac{1}{\tau B_n^3} \sum_{k=1}^{n} \int_{-k}^{k} |x|^3 \frac{1}{2k} \mathrm{d}x$$

$$= \frac{1}{4\tau B_n^3} \sum_{k=1}^{n} k^3$$

$$= \frac{1}{4\tau} \cdot \frac{18^{\frac{3}{2}}}{[n(n+1)(2n+1)]^{\frac{3}{2}}} \cdot \frac{n^2(n+1)^2}{4}$$

$$= \frac{18^{\frac{3}{2}}}{16\tau} \cdot \frac{n(n+1)}{(2n+1)\sqrt{n(n+1)(2n+1)}} \longrightarrow 0 \quad (n \to \infty)$$

林德伯格条件成立, 故 $\{X_k\}$ 适用中心极限定理.

【评注】《概率论基础》定理 5.5.3 也可用于本题, 可见它是一种便于使用的中心极限定理类型.

57. 试问对下列独立随机变量序列, 李雅普诺夫定理是否成立?

(1) X_k:

	$-\sqrt{k}$	$\sqrt{k}$
	$\frac{1}{2}$	$\frac{1}{2}$

(2) X_k:

	$-k^\alpha$	0	k^α
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

, $\alpha>0$

解 (1)

$$EX_k = 0, \quad DX_k = k$$

$$B_n^2 = \sum_{k=1}^{n} k = \frac{1}{2} n(n+1)$$

取 $\delta = 2$, 则

$$\frac{1}{B_n^{2+2}} \sum_{k=1}^{n} E|X_k|^{2+2} = \frac{4n(n+1)(2n+1)}{6n^2(n+1)^2} \longrightarrow 0 \quad (n \to \infty)$$

所以李雅普诺夫定理成立.

(2)

$$EX_k = 0, \quad DX_k = \frac{2}{3} k^{2\alpha}$$

又函数 $x^{2\alpha}$ 和 $(x+1)^{3\alpha}$ 单调上升, 在 $x\in[k-1,k]$ 有 $k^{2\alpha}\geqslant x^{2\alpha}$, $k^{3\alpha}\leqslant(x+1)^{3\alpha}$, 于是

$$\sum_{k=1}^{n}k^{2\alpha}=\sum_{k=1}^{n}\int_{k-1}^{k}k^{2\alpha}\mathrm{d}x\geqslant\sum_{k=1}^{n}\int_{k-1}^{k}x^{2\alpha}\mathrm{d}x=\int_{0}^{n}x^{2\alpha}\mathrm{d}x$$

$$\sum_{k=1}^{n}k^{3\alpha}\leqslant\sum_{k=1}^{n}\int_{k-1}^{k}(x+1)^{3\alpha}\mathrm{d}x=\sum_{k=1}^{n}\int_{k}^{k+1}u^{3\alpha}\mathrm{d}u=\int_{1}^{n+1}x^{3\alpha}\mathrm{d}x$$

这样

$$B_n^2=\frac{2}{3}\sum_{k=1}^{n}k^{2\alpha}\geqslant\frac{2}{3(2\alpha+1)}n^{2\alpha+1}$$

取 $\delta=1$, 得

$$\begin{aligned}\frac{1}{B_n^{2+1}}\sum_{k=1}^{n}E|X_k|^{2+1}&=\frac{1}{B_n^3}\sum_{k=1}^{n}\frac{2}{3}k^{3\alpha}\\&\leqslant\frac{2}{3B_n^3}\int_{1}^{n+1}x^{3\alpha}\mathrm{d}x\\&\leqslant\frac{\sqrt{3}(2\alpha+1)^{\frac{3}{2}}}{\sqrt{2}\sqrt{n^{3(2\alpha+1)}}}\left(\frac{(n+1)^{3\alpha+1}}{3\alpha+1}-\frac{1}{3\alpha+1}\right)\\&\longrightarrow 0\quad(n\to\infty)\end{aligned}$$

同样李雅普诺夫定理成立.

【评注】李雅普诺夫条件是历史上最早找到的一个使中心极限定理成立的便于使用的充分条件, 至今仍有实用价值.

*58. 求证: 当 $n\to\infty$ 时

$$\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_{0}^{1+t\sqrt{\frac{2}{n}}}z^{\frac{n}{2}-1}\mathrm{e}^{-\frac{nz}{2}}\mathrm{d}z\ \to\ \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}\mathrm{e}^{-\frac{z^2}{2}}\mathrm{d}z$$

证 设独立同分布随机变量序列 $\{\xi_n\}$ 均服从 Γ 分布 $\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$, 则 $E\xi_1=1$, $D\xi_1=2$.

由 Γ 分布的再生性知 $\eta_n = \sum\limits_{k=1}^{n} \xi_k$ 服从 Γ 分布 $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, 且 $E\eta_n = n$, $D\eta_n = 2n$. 由林德伯格-莱维中心极限定理,

$$P\left\{\frac{\eta_n - n}{\sqrt{2n}} < t\right\} = P\{\eta_n < n + t\sqrt{2n}\}$$

$$= \int_0^{n+t\sqrt{2n}} \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}} \mathrm{d}x$$

令 $x = nz$,

$$P\left\{\frac{\eta_n - n}{\sqrt{2n}} < t\right\} = \frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_0^{1+t\sqrt{\frac{2}{n}}} z^{\frac{n}{2}-1} \mathrm{e}^{-\frac{nz}{2}} \mathrm{d}z$$

$$\longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x \quad (n \to \infty)$$

【评注】用概率论方法证明一个分析结果. 注意到 $\eta_n \sim \chi_n^2$, 因此本题等价于 χ_n^2 分布渐近正态分布.

*59. 独立随机变量序列 $\{\xi_k\}$, 对一切 k, ξ_k 以概率 $\frac{1}{2}$ 分别取值 $\pm k^s$,

(1) 试证当 $s < \frac{1}{2}$ 时, 大数定律成立;

(2) 试利用中心极限定理证明: 当 $s \geqslant \frac{1}{2}$ 时, 大数定律不成立.

证 (1) $E\xi_k = 0$, $D\xi_k = k^{2s}$, 当 $s < \frac{1}{2}$ 时计算得

$$\frac{1}{n^2} D\left(\sum_{k=1}^{n} \xi_k\right) = \frac{1}{n^2} \sum_{k=1}^{n} k^{2s} \leqslant \frac{n^{2s} n}{n^2} = n^{2s-1} \longrightarrow 0 \quad (n \to \infty)$$

满足马尔可夫条件, 所以大数定律成立.

(2) 对 $\delta > 0$,

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E|\xi_k-a_k|^{2+\delta}=\frac{\sum\limits_{k=1}^{n}k^{s(2+\delta)}}{\left(\sum\limits_{k=1}^{n}k^{2s}\right)^{\frac{2+\delta}{2}}}$$

$$\approx\frac{n^{s(2+\delta)+1}}{\left(n^{2s+1}\right)^{1+\frac{\delta}{2}}}$$

$$=\frac{n^{2s+s\delta+1}}{n^{2s+s\delta+\frac{\delta}{2}+1}}$$

$$=\frac{1}{n^{\frac{\delta}{2}}}\longrightarrow 0$$

即 $\{\xi_k\}$ 对任意 s 满足李雅普诺夫条件, 故对它总是成立中心极限定理.

在 $s\geqslant\frac{1}{2}$ 时,

$$a_k=E\xi_k=0,\quad B_n^2=\sum_{k=1}^{n}D\xi_k=\sum_{k=1}^{n}k^{2s}$$

对任意 $\varepsilon>0$, 有

$$\frac{\varepsilon^2n^2}{B_n^2}=\frac{\varepsilon^2n^2}{\sum\limits_{k=1}^{n}k^{2s}}\leqslant\frac{\varepsilon^2n^2}{\sum\limits_{k=1}^{n}k}=\frac{2\varepsilon^2n^2}{n(n+1)}\longrightarrow 2\varepsilon^2\quad(n\to\infty)$$

这样当 n 充分大时, $\frac{\varepsilon n}{B_n}<\sqrt{3}\varepsilon$, 且有

$$P\left\{\frac{1}{n}\left|\sum_{k=1}^{n}\xi_k\right|<\varepsilon\right\}=P\left\{\frac{B_n}{n}\cdot\frac{1}{B_n}\left|\sum_{k=1}^{n}\xi_k\right|<\varepsilon\right\}$$

$$=P\left\{\frac{1}{B_n}\left|\sum_{k=1}^{n}\xi_k\right|<\frac{\varepsilon n}{B_n}\right\}$$

$$\leqslant P\left\{\frac{1}{B_n}\left|\sum_{k=1}^{n}\xi_k\right|<\sqrt{3}\varepsilon\right\}$$

因而,

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\left|\sum_{k=1}^{n}\xi_k\right| < \varepsilon\right\} \leqslant \int_{-\sqrt{3}\varepsilon}^{\sqrt{3}\varepsilon} \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = A < 1$$

所以, $s \geqslant \dfrac{1}{2}$ 时对 $\{\xi_k\}$ 大数定律不成立.

【评注】大数定律能否成立, 矩或尾端概率起重大作用, 概率分布

	$-k^s$	k^s
	$\frac{1}{2}$	$\frac{1}{2}$

是一个典型例子, $s = \dfrac{1}{2}$ 是成立与否的临界值, 仅对比它小的 s 值才成立. 据此可见对本章习题 5 成立, 习题 6 (1) 不成立.

本题 (2) 的证明思路与习题五 51 题紧密相关, 当 $2s \geqslant 1$ 时, $B_n^2 = \sum\limits_{k=1}^{n} k^{2s} = o\left(n^2\right)$ 不成立.

*60. 用概率论方法证明: 当 $n \to \infty$ 时, $\mathrm{e}^{-n}\sum\limits_{k=0}^{n}\dfrac{n^k}{k!} \to \dfrac{1}{2}$.

证 设有独立同分布随机变量序列 $\{\xi_k\}$ 服从 $\lambda = 1$ 的泊松分布, 由泊松分布的再生性知 $\eta_n = \sum\limits_{k=1}^{n}\xi_k$ 服从参数 $\lambda = n$ 的泊松分布, 即

$$P\{\eta_n = k\} = \frac{n^k}{k!}\,\mathrm{e}^{-n},\ (k = 0, 1, 2, \cdots),\ 且\ E\eta_n = D\eta_n = n$$

由林德伯格–莱维中心极限定理,

$$P\left\{\frac{\eta_n - n}{\sqrt{n}} \leqslant 0\right\} = P\{\eta_n \leqslant n\} = \mathrm{e}^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$$

$$\longrightarrow \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x = \frac{1}{2}$$

【评注】用概率论方法证明分析结果的又一有趣的例子.

习题总评

第五章理论性较强, 因此习题的选择主要是配合正文. 许多题目是为加深对正文的理解而配置的. 有些提供正、反例证, 有些提供应用题, 有些则作了解释或补充. 不但对主要结果而且还对主要方法设置相应的题目.

用切比雪夫型不等式证明大数定律和用特征函数法证明中心极限定理是两大主题. 不少题目是对导出的一些充分条件的验证. 各种收敛性也带来一大批题目.

关于极限定理的计算题, 本质上只有那么几种类型, 希望做完之后自己做个小结.

章后小议

三段式展开的结构, 当然多花时间, 因为前者是后者的简单特例, 证得后者立即推出前者, 从一般到特殊的讲法肯定可以节省时间. 但是三段式的结构也不浪费时间, 因为学到更多知识, 了解更多历史, 有助于拓宽眼界.

伯努利大数定律原始的证明是纯分析的, 在 Uspensky 的书中有介绍. 《概率论基础》中它是作为切比雪夫大数定律的特例推出的, 立刻又用切比雪夫不等式直接证明. 泊松大数定律 (它是独立而不同分布的代表) 也类似处理. 在矩法发明后, 当然没有必要在教本中写出繁复的分析证明. 但棣莫弗–拉普拉斯定理在书中是用斯特林公式直接证明的, 得到了 "副产品" 局部极限定理, 它有许多应用, 也存历史原貌.

独立同分布场合极限定理的证明，一用特征函数就相当简单明了. 读者将在此场合体会证法的要点，为以后讨论更复杂的场合作好准备.

对上述两个场合所得结果的有关应用，特别是结合数理统计，新版作了较充分的介绍. 于是读者对概率论成为统计学的理论基础，将有更深的认识.

《概率论基础》第五章中的收敛性一节，完成了两大任务. 一是证明了特征函数与分布函数对应的连续性定理，为特征函数成为研究概率极限定理主要工具打通了最后一道关卡. 二是提出随机变量的四种收敛性，为拓广概率论的研究领域摆开阵势. 作为"副产品"，顺带证明了波赫纳尔-辛钦定理与赫格洛茨定理，前者给出特征函数的充要条件，后者在平稳随机过程论中起重要作用. 最后，新版给出帕斯卡分布逼近埃尔朗分布的一个证明，给等待时间分布对比作一个小结.

多元中心极限定理经常被引用，但少见证明. 本书巧用前期积累的知识，轻松给出.

承继复旦教科书的传统，《概率论基础》对连续性定理及波赫纳尔-辛钦定理写出详细的证明，这部分内容应由任课教师自主取舍. 最后两节也是如此，或许博雷尔-康特立引理及博雷尔强大数定律应列入讲课大纲.

本章讨论的概率极限定理可列表分类如下:

	大数定律	强大数定律	中心极限定理
伯努利试验	伯努利~	博雷尔~	棣莫弗-拉普拉斯~
独立同分布	辛钦~	科尔莫戈罗夫~	林德伯格-莱维~
一般场合	泊松~ 切比雪夫~ 马尔可夫~	科尔莫戈罗夫~	林德伯格-费勒~

教学札记之二十五

统计学和概率论

一、数理统计的理论框架

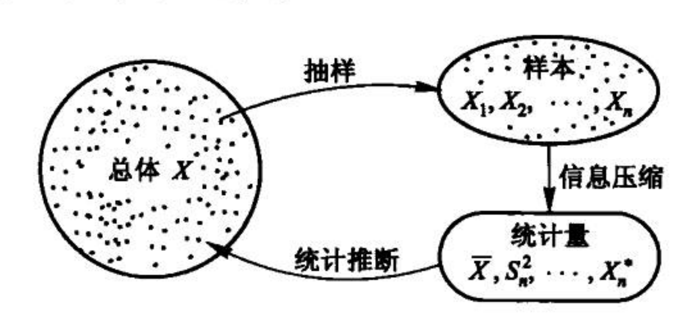


图 5-2

一个典型的统计问题可描述如下: 进行一系列随机实验, 从实验中收集到数据, 任务是从这些数据中提取信息, 解释结果, 并作出某些结论.

所考察对象全体称为**总体**. 例如它可以是全市收看有线电视的家庭全体, 也可以是某种电子元件的使用寿命, 这个总体有时还可以是想象中的, 例如使用某种抗癌新药的潜在患者全体.

一般对总体关心的是一个数量指标, 例如, 某频道收视与否, 寿命, 五年存活率等, 因此通常用一个随机变量 X 来等同一个总体. 若关心的是多个指标, 则归入**多元分析**这一统计分支, 此处不论.

为了了解总体的这个指标, 可以对总体中的每个个体作测量, 但或由于总体太大力所不能及或花费太多, 或测量带有破坏性, 或带有潜在性, 通常抽样是唯一可行的方法. 这时得到总体的一个**样本**. 统计学的任务是通过这个样本 $X_1, X_2, \cdots, X_n$ 来对总体 X 作出种种论断, 称之为**统计推断**, 有时则还作出某种**决策**, 这时还得考虑采取相应行动的后果.

统计学中假定样本依某种随机方式从总体中抽取以保证它对总体有某种代表性, 从而使推断有意义, 这时概率论成为统计方法

的理论依据.

这样一来, 统计学成为一门归纳的学科, 它从某些经验事实出发, 对总体作出结论, 即从部分推断总体. 但这正是一切实验科学所必经之路, 因此统计学成为从实验数据到构建理论模型的主要科学工具, 这就决定了统计学在现代科学体系中的重要地位.

相反地, 概率论则是一门演绎的科学, 它从某些公理化的事实出发, 利用形式逻辑作推理, 建立自己的理论体系.

概率论与数理统计就是这样相辅相成: 概率论直接用于实际, 基本上都要通过数理统计; 而数理统计的理论基础则完全建立在概率论之上.

为了做到这点, 当代统计学采用如下框架:

设 $X \sim F(x,\theta)$, 其中 $F(x,\theta)$ 称为**总体 X 的分布**, θ 称为**参数**, $F(x,\theta)$ 至少部分未知.

$X_1, X_2, \cdots, X_n$ 是相互独立, 相同分布的随机变量, 同服从分布 $F(x,\theta)$, 称为**容量**为 n 的**简单随机样本**, 简称**样本**.

实际操作中观察到的是 n 个观察值 $x_1, x_2, \cdots, x_n$, 当代统计中把它们看作独立同分布的随机变量的观察值, 这是使原本的统计学变成数理统计学的关键, 只有这样才能对各种统计方法施行的效果进行概率的评价, 从而使统计学成为科学, 成为一般科学的方法论.

这里采用独立同分布的假定事实上是数理统计建立初期的原型, 便于建立系统的理论. 它把历史上两大类统计问题统一起来, 功劳很大.

【概率 p 的估计】 统计学历史上的一大类问题是关于事件 A 发生率 $p = P(A)$ 的估计. 从伯努利建立伯努利概型起, 便涉及这类问题.

从概率论角度看, 已知 p 之后, 要求研究 A 出现的频率, 即在 n 次独立重复试验中 A 出现 k 次的概率, 得到二项分布 $\binom{n}{k} p^k q^{n-k}$,

$p+q=1, k=0,1,2,\cdots,n$. 对频率与概率关系的研究导致概率论两大定律, 即大数定律和中心极限定理的发现.

另一方面,若在 n 次独立重复试验中 A 共出现 k 次, 如何来估值 p 呢? 从雅科布 · 伯努利到丹尼尔 · 伯努利, 到贝叶斯, 再到拉普拉斯都讨论过这个问题, 而拉普拉斯更把该问题与法国人口的抽样调查结合在一起.

至于英国的统计学先驱格朗特 (J. Graunt, 1620 — 1674), 哈雷 (Halley, 1656 — 1742) 和比利时的凯特勒 (L. Quetelet, 1796 — 1874) 也都是从人口统计学开始发展统计理论的.

因此历史上人口统计 (人口总数统计, 新生婴儿性别比统计, 寿命统计) 的实际需要, 推动对伯努利试验中概率 p 的估计的研究.

【测量误差的估计】 早期概率论大家如拉普拉斯, 高斯等也都是天文学家, 因此对天文和测地中观察误差的研究一直是他们关心的重要课题. 这些研究使得拉普拉斯与高斯发现了正态分布. 发现误差分布有规律存在最初是英国的辛普森 (T. Simpson, 1710 — 1761) 于 1755 年在研究算术平均值时发现的. 后来这个课题的研究一直持续不断. 若 μ 是真值, ε 是误差, 则 $X_i=\mu+\varepsilon_i, i=1,2,\cdots,n$ 是 n 次独立观察, $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 服从相同的对称分布, 它们可以是 $[-a,a]$ 均匀分布、辛普森 $[-a,a]$ 三角形分布、拉普拉斯分布或正态分布, 任务是估计 μ, 并估计观察误差.

这两个古典统计问题形式十分不同: 从今天看, 一个涉及离散型, 一个涉及连续型; 一个面对有限总体人口已有观察数据在手, 一个面对无穷总体, 可设计潜在无限次的观察.

但是在上述框架下, 它们被统一起来.

在人口统计中, 只要设 X 服从概率为 p 的伯努利分布, 而 $X_1,\cdots,X_n$ 是它的 n 个独立观察, 便可纳入模型, 这相当于要求抽样是有放回的. 对于人口总数很大, 抽样量相对很小, 这个假定至少是近似成立的.

当然随着统计理论和实践的进一步发展，抽样方法在社会生活中的应用领域日增，方式多样，对有限总体的抽样一般都是不放回的，而是要求每个个体有同等机会被选入样本. 这种抽样方式保持了同分布的要求但不再具有独立性. 一般把这种样本亦称为简单随机样本. 对这种随机样本的研究形成统计学的一门分支，称为**抽样调查**，它已突破上述框架，但吸收该框架的大量研究成果，成为在社会生活中广泛使用的一种统计技术，在《概率论基础》第四章§2 例 10 中略有涉及.

二、数据的整理与压缩、统计量

统计学是研究收集和分析数据的学问，一般更着重于分析数据方面. 面对一个容量为 n 的样本，如何利用它研究总体的方方面面呢? 第一步一定是整理数据. 历史上，统计学中有一个学派对此作了系统的研究，这些大体构成如今描述性统计的内容. 还有，随着计算机的发展，有了强大的计算能力，图形功能与交互式会话的能力，以数据分析为名的一些新方法也不断呈现，例如直方图，枝叶图，箱图，经验分布函数，曲线拟合技术等图形技术和各种汇总指标的计算. 总的说，这些发展与概率论中对分布描述的发展是相互平行相互推动的.

统计量概念的提出是最具重要性的，所谓**统计量**是样本的函数 $f(X_1, X_2, \cdots, X_n)$，它由样本完全确定，不含未知参数. 提出统计量的目的是压缩数据即提炼信息，因为面对大的容量 n，一大堆数据 $x_1, x_2, \cdots, x_n$ 显示不了什么，需要加以精炼，集中，变成少数几个量，这当中又希望不要丢掉有用的信息. 历史上研究得最早最多的是关于样本算术平均值 $\overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$，后来又研究中位数 X_{med} 和样本方差 $S_n^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$. 最基本的可归纳为两大类: 一

类是矩统计量 $A_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$, 一类是顺序统计量, 包括极大值 X_n^* 和极小值 X_1^* 以及中位数, 分位数等. 然后利用它们再根据问题的需要构造出种种统计量.

统计量是样本的函数, 是随机变量, 因此求它们的分布是最根本的事情. 到此, 我们就可以明白为什么基础概率论中要花那么大的力气来讨论求随机变量函数的分布. 事实上, 主要是为了适应统计学的需要.

统计量的分布常称为**抽样分布**, 因此统计学讲完基本概念后的第一章总是讲统计量的抽样分析. 在《概率论基础》中讲统计三大分布 χ^2, F, t 分布即为此. 另外, 讨论了极大值与极小值的联合分布以及极差的分布也为此.

能求出精确分布的统计量是极少数, 大量的只能求渐近分布, 这也要依靠概率论, 主要是利用中心极限定理, 因此在统计学中正态分布仍是最重要的分布.

为压缩数据而不丢掉有用信息提出的主要概念是 "充分性", 这是联系总体分布与参数的一个重要性质, 指数族分布 (见习题三 *11 题) 的提出与此相关.

三、三类统计推断问题

统计学从部分推断整体, 属于归纳推理. 经长期发展, 最后发展成三类问题, 它们一起构成推断统计学的主要部分. 推断统计学是与描述统计学平行的另一类统计方法与理论, 一般认为它更代表近代统计学的真谛.

第一类问题是**估计理论**, 最基本的是未知参数 θ 的估计. 一方面提出一些系统性的估计方法, 主要有矩法, 最大似然估计法和最小二乘法三种, 在《概率论基础》都有涉及; 另一方面是提出一些筛选标准及优良性准则, 最重要的有无偏性, 有效性和相合性三个, 《概率论基础》也都讲到.

第二类问题是**假设检验**. 这与利用试验资料建立科学模型有关. 在统计学中发展成最主要的一类推断方法. 这类方法密切联系于检验统计量的选取, 而求检验量的分布更是重中之重, 与概率论关系密切.《概率论基础》中在第二章 §4 例 2 及 (n,c) 方案一例中作了简单的介绍, 而第五章 §3 例 3 则指出大多数场合只能用渐近分布, 因而中心极限定理起重要作用.

第三类问题是**置信区间**, 给出另一种形式的估计——区间估计.《概率论基础》给出仅有的一例, 见 (5.1.35) 式.

统计学在 19 世纪末到 20 世纪初由高尔顿 (F. Galton, 1822—1911), 卡尔・皮尔逊 (Karl Pearson, 1857—1936) 主导, 20 世纪 20 年代前后由费希尔 (R. A. Fisher, 1890—1962) 大加发展, 20 世纪 30 年代后期奈曼 - E. 皮尔逊创立假设检验新理论成为新的权威.

瓦尔德 (A. Wald, 1902—1950) 在 20世纪 50 年代初曾以统计决策函数之名把这几类问题统一处理.

四、正态总体下的统计理论

当代统计学的典范是在总体服从正态分布 $N(\mu,\sigma^2)$ 的情况下建立起来的. 这个模型, 一方面由于正态分布是实际问题中最常见的, 有中心极限定理作理论依托, 特别在测量误差这类古典问题中还有高斯正态性推导的有力背景; 另一方面是由于在正态性假定下, 许多精确分布能求得, 这集中体现在下列费希尔引理之中:

费希尔引理 若 $X_1, X_2, \cdots, X_n$ 相互独立, 均服从 $N(\mu,\sigma^2)$, 记

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

则 (1) $\overline{X}$与 S_n^2 相互独立,

(2) $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$,

(3) $\dfrac{nS_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

这个定理是习题四 **60 题, 它给出 $\overline{X}$ 与 S_n^2 的分布, 它们又相互独立, 因而 $(\overline{X}, S_n^2)$ 的联合分布容易写出. 这种场合容易导出费希尔 F 统计量与哥塞特 (Gosset, 1876－1937) t 统计量的分布, 从而能进行一系列统计推断活动. 在这种特殊场合建立了近代统计学的漂亮而实用的理论.

关于 χ^2 分布, F 分布与 t 分布的推导见《概率论基础》正文, 亦见教学札记之二十六 "统计学三大分布的推导".

这部分理论的建立完全依托概率论, 概率论这部分的内容也因统计学的发展而充实.

五、统计学概观

正态总体场合利用正态分布的优良性质建立起来的优美统计理论成为特例, 其他分布无法企及, 稍微满意的是关于比例 p 的统计, 它一般用频率 $\frac{\mu_n}{n}$ 来估计, 而 μ_n 服从二项分布, 这时三大统计推断问题能得到比较满意的解决, 此外关于泊松总体, Γ 分布总体对均值的统计推断比较好办, 对样本方差就难.

因此在 20 世纪 30 年代就发展出一种新的方法, 称为**非参数统计**, 目的主要是去掉正态性假定代之以连续分布之类的假定, 取得了一定的进展, 形成非参数这一分支.

另外一个分支是**相关分析**与**回归分析**, 都与高尔顿关于生物统计的研究有关, 特别是回归分析已发展成为统计学中最常用的方法之一. 应当说, 这部分的研究是统计学带动了概率论, 包括带动多元正态分布理论的研究. 《概率论基础》反映这部分内容的是第四章 §2 关于条件期望, 最佳线性预测一小节.

费希尔建立的**试验设计**与**方差分析**方法也成为统计学有广泛应用的分支之一, 在 20 世纪中期, 它们与回归分析等统一成**线性模型**, 后期又发展成**广义线性模型**, 这类模型由于有统计软件包的支持已很有实用价值.

多元分析在 20 世纪二、三十年代由费希尔学派发展时只是一

些理论模型, 如今在计算机软件支持下已成为很强有力的分析工具.

计算机的发展在各个方面改变了统计学的面貌. 总的说, 20 世纪后期统计学的发展主要体现在**计算统计**上, 各种与计算有关的方法被提出, 例如刀切法、自助法、交叉核实法、模型的选择与诊断、函数的拟合、EM 算法、MCMC 等, 对稳健性也提出更高的要求, 未来的发展也必与处理大数据集相关.

去掉样本中的独立性假定就进入**时间序列分析**, 其理论基础也从概率论进入随机过程, 一般要引进平稳性假定, 从而与平稳随机过程紧密相关. 这类过程前有科尔莫戈罗夫的理论成果后有维纳的研究, 近代已由统计学家改进成程式化的处理方法, 获得广泛应用, 如博克斯–詹金斯 (Box - Jenkins) 模型. 另一发展方向是由控制论专家卡尔曼 (Kalman) 完成, 用控制论中传统的运动方程和观测方程发展成多维递推模式, 也得到成功应用. 平稳过程中的谱分析理论在时间序列分析中也有相应发展, 较多用于电信系统, 也有用于经济分析.

统计学向各应用领域的渗透形成新学科成为 20 世纪的一道风景线, 由 "生物+统计" 而形成的新词 (也即新学科) Biometrika 之后有 "Econometrics", "Technometrics" $\cdots$, 这类词已达数十个.

教学札记之二十六

统计学三大分布的推导

统计学三大分布, χ^2 分布, F 分布, t 分布, 由于它们的重要性, 已提出诸多推导方法. 这些是基础概率论中求随机变量函数的分布方法之范例, 特举要介绍.

一、χ^2 分布

若 $Z_1, \cdots, Z_n$ 相互独立, 均服从标准正态分布 $N(0,1)$, 记

$\chi_n^2 = Z_1^2 + \cdots + Z_n^2$, 定义 χ_n^2 的分布为 χ^2 分布, 其概率密度函数表达式为

$$p(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}}, \quad x > 0 \tag{1}$$

其中 n 称为自由度或参数. 通常这个分布也记作 χ_n^2 或 $\chi^2(n)$, 并不会造成混淆.

χ^2 分布于 1900 年由卡尔 · 皮尔逊再引进, 是统计学中最重要的分布. 后发现它可以归入 Γ 分布, 为 $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

求 (也是 "记忆") χ_n^2 分布的数学期望和方差的最好方式是记住如下推导:

显然

$$EZ_i = 0, \quad DZ_i = EZ_i^2 = 1, \quad DZ_i^2 = EZ_i^4 - (EZ_i^2)^2 = 3 - 1 = 2$$

故

$$E\chi_n^2 = EZ_1^2 + \cdots + EZ_n^2 = n$$
$$D\chi_n^2 = DZ_1^2 + \cdots + DZ_n^2 = 2n$$

由习题五 *37 题及习题五 *58 题知, $\frac{\chi_n^2 - n}{\sqrt{2n}}$ 渐近 $N(0,1)$, 即 χ_n^2 以正态分布为渐近分布.

χ^2 分布的推导方法很多, 以下是最典型的.

1. 从定义 $\chi_n^2 = Z_1^2 + \cdots + Z_n^2$ 出发, 用 n 重积分直接计算, 化为 n 维球坐标, 最后导出表达式. 请参看习题三 *37 题及其解答.

2. 直接计算 χ_1^2 的表达式, 再用 χ^2 分布的可加性, 用归纳法导出.

《概率论基础》在第三章 § 3 导出 χ_1^2 的表达式 (3.3.10), 并在例 7 证明了 χ^2 分布的可加性, 指明了这种推导法.

3. 借助 χ^2 分布与 Γ 分布的关系, 由 χ_n^2 分布为 $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ 立即得出.

4. 用特征函数法. χ_n^2 的特征函数 $(1 - 2\,\mathrm{i}t)^{-\frac{n}{2}}$ 与密度函数 (1) 对应, 这里隐藏着许多推导法.

二、χ 分布

利用这个机会, 我们来推导 $\chi_n = \sqrt{\chi_n^2} = \sqrt{Z_1^2 + \cdots + Z_n^2}$ 的分布密度函数, 由于它在 n 维欧几里得空间中表示 $(Z_1, \cdots, Z_n)$ 与原点的距离, 因此这个分布在自然科学中有许多应用.

显然, 对 $x \leqslant 0$,

$$F_{\chi_n}(x) = P\{\chi_n < x\} = 0$$

而对 $x > 0$,

$$\begin{aligned} F_{\chi_n}(x) &= P\{\chi_n < x\} = P\{\chi_n^2 < x^2\} \\ &= \int_0^{x^2} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} \mathrm{e}^{-\frac{y}{2}} \, \mathrm{d}y \end{aligned}$$

$$\begin{aligned} p_{\chi_n}(x) = F'_{\chi_n}(x) &= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} (x^2)^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x^2}{2}} \cdot 2x \\ &= \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} x^{n-1} \mathrm{e}^{-\frac{x^2}{2}}, \quad x > 0 \end{aligned} \tag{2}$$

注意当 $n = 2$ 时, $\rho = \sqrt{Z_1^2 + Z_2^2}$ 的分布即为瑞利分布 (3.3.41),

$$R(r) = r\,\mathrm{e}^{-\frac{r^2}{2}}, \quad r > 0$$

当 $n = 3$ 时, $S = \sigma\sqrt{Z_1^2 + Z_2^2 + Z_3^2}$ 的分布即为麦克斯韦分布律,

$$p(s) = \sqrt{\frac{2}{\pi}} \frac{s^2}{\sigma^3} \mathrm{e}^{-\frac{s^2}{2\sigma^2}}, \quad s > 0$$

见习题三 36 题.

顺便指出 $\dfrac{\chi_n}{\sqrt{n}}$ 的密度函数为

$$p_{\frac{\chi_n}{\sqrt{n}}}(y) = \frac{n^{\frac{n}{2}} y^{n-1}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} \mathrm{e}^{-\frac{ny^2}{2}}, \quad y > 0 \tag{3}$$

这在后面用到.

三、F 分布

若 $X \sim \chi_m^2, Y \sim \chi_n^2$ 且 X 与 Y 独立, 则称 $F = \dfrac{X/m}{Y/n}$ 服从 F 分布, 其密度函数为

$$p(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}}, \quad x > 0 \tag{4}$$

其中参数 m, n 分别称为分子自由度与分母自由度, 这分布记作 $F(m,n)$. 显然 $\dfrac{1}{F} \sim F(n,m)$.

在处理 χ_n^2 变量时通常都除以它的自由度, 大体可理解为关心的是每个自由度的变差贡献率, 这样来定义统计量 F 最早由美国统计学家斯内德克 (Snedecor) 引进.

费希尔引进了 F 分布的一种变形, 二者事实上是等价的, 目前用的较多的形式是 F 分布. 这个分布取名 F, 正是为了表彰费希尔的贡献.

F 分布的推导相对比较单纯, 主要有下列二法:

1. 用变换法. 作变换 $S = X + Y$, $F = \dfrac{X/m}{Y/n}$, 利用 χ^2 分布的可加性立得 $S \sim \chi_{m+n}^2$. 这时用 Γ 分布的刻画性质知 S 与 F 独立, 因此很容易得到 F 的密度表达式, 在《概率论基础》第三章 §3 例 7 中就是这样做的.

2. 用除法公式.

$$\begin{aligned} p_F(z) &= \int_{-\infty}^{+\infty} |v| p_{\frac{X}{m}}(zv) p_{\frac{Y}{n}}(v)\, \mathrm{d}v \\ &= \int_0^{+\infty} |v| \frac{m^{\frac{m}{2}}}{2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)} (zv)^{\frac{m}{2}-1} \mathrm{e}^{-\frac{mzv}{2}} \cdot \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} v^{\frac{n}{2}-1} \mathrm{e}^{-\frac{nv}{2}}\, \mathrm{d}v \end{aligned}$$

$$
= \frac{m^{\frac{m}{2}} n^{\frac{n}{2}}}{2^{\frac{m+n}{2}} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} z^{\frac{m}{2}-1} v^{\frac{m+n}{2}-1} \mathrm{e}^{-\frac{mz+n}{2} v} \mathrm{d}v
$$

$$
= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}} z^{\frac{m}{2}-1}}{\left(1+\frac{m}{n} z\right)^{\frac{m+n}{2}}}, \quad z \geqslant 0
$$

四、t 分布

t 分布一般采用便于统计学中应用的形式定义.

设 $X \sim N(0,1)$, $Y \sim \chi_n^2$, 且 X 与 Y 相互独立, 则 $T = \dfrac{X}{\sqrt{Y/n}} \sim t(n)$ 分布, 其密度函数的表达式为

$$
p(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < +\infty \qquad (5)
$$

参数 n 亦称为自由度.

t 分布关于 $x=0$ 对称, 属于肥尾分布, 因此只存在 $r<n$ 阶矩. 特别当 $n=1$ 时它化为柯西分布, 连数学期望都不存在.

t 分布于 1908 年由英国统计学家哥塞特在以 "Student" 为笔名的一篇论文中引进, 开创了小样本理论这一分支, 是统计学中三大分布之一, 自由度一词也是哥塞特最早使用的.

$t(n)$ 以正态分布为极限, 事实上当 n 较大时, 二者已很接近.

t 分布的推导法也很多, 下面列举其中重要的几种:

1. 用变换法. 例如在《概率论基础》第三章§3 例 8 中通过引进增补变量 $S=Y$, 建立 (X,Y) 与 (S,T) 的一一变换, 从而得到 (S,T) 的联合密度函数, 再通过求边际密度得到 T 的密度函数 (5).

2. 用除法公式, 即《概率论基础》(3.3.23). 利用 X 与 Y 的

独立性, $X \sim N(0,1)$ 而 $\sqrt{Y/n}$ 的密度由 (3) 表出, 故

$$\begin{aligned} p_T(z) &= \int_{-\infty}^{+\infty} |v| p_X(zv) p_{\sqrt{Y/n}}(v)\,\mathrm{d}v \\ &= \int_0^{+\infty} v \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2v^2}{2}} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} v^{n-1} \mathrm{e}^{-\frac{nv^2}{2}}\,\mathrm{d}v \\ &= \frac{1}{\sqrt{2\pi}} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} \int_0^{+\infty} v^n \mathrm{e}^{-\frac{(n+z^2)v^2}{2}}\,\mathrm{d}v \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}} \end{aligned}$$

3. 化为极坐标. 作变换

$$x = r\cos\theta, \quad \sqrt{y} = r\sin\theta, \quad (r>0,\ 0<\theta<\pi)$$

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ 2r\sin^2\theta & 2r^2\sin\theta\cos\theta \end{vmatrix} = 2r^2\sin\theta$$

则

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} \mathrm{e}^{-\frac{y}{2}}\,\mathrm{d}x\mathrm{d}y \\ &\qquad = \frac{1}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)2^{\frac{n-1}{2}}} r^n \sin^{n-1}\theta\, \mathrm{e}^{-\frac{r^2}{2}}\,\mathrm{d}r\mathrm{d}\theta \qquad (*) \end{aligned}$$

这时 r 与 θ 已分离, 再作变换

$$t = \frac{x}{\sqrt{y/n}} = \sqrt{n}\cot\theta$$

因而有 $\mathrm{d}t = -\dfrac{\sqrt{n}}{\sin^2\theta}\,\mathrm{d}\theta$

$$\sin^2\theta = \frac{1}{1+\cot^2\theta} = \frac{1}{1+\dfrac{t^2}{n}}$$

注意变换行列式应为正, (*) 化为

$$\frac{1}{\sqrt{n\pi}\Gamma\left(\dfrac{n}{2}\right)2^{\frac{n-1}{2}}}\frac{r^n \mathrm{e}^{-\frac{r^2}{2}}}{\left(1+\dfrac{t^2}{n}\right)^{\frac{n+1}{2}}}\,\mathrm{d}t\mathrm{d}r \qquad (**)$$

对 r 求积得到 $t(n)$ 的密度表达式 (5).

4. 利用 F 分布导出. 注意到 $T^2 = \dfrac{X^2}{Y/n}$, 后者显然服从 $F(1,n)$, 这提供了从 F 分布导出 t 分布的一条途径.

显然 T 是对称的, 因此

$$G(x) = P\{0 < T < x\} = \frac{1}{2}P\{T^2 < x^2\}$$

$$= \frac{1}{2}\int_0^{x^2} \frac{\Gamma\left(\dfrac{1+n}{2}\right)\left(\dfrac{1}{n}\right)^{\frac{1}{2}} z^{\frac{1}{2}-1}}{\Gamma\left(\dfrac{1}{2}\right)\Gamma\left(\dfrac{n}{2}\right)\left(1+\dfrac{1}{n}z\right)^{\frac{1+n}{2}}}\,\mathrm{d}z$$

$$G'(x) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)\left(\dfrac{1}{n}\right)^{\frac{1}{2}}(x^2)^{\frac{1}{2}-1}}{\Gamma\left(\dfrac{1}{2}\right)\Gamma\left(\dfrac{n}{2}\right)\left(1+\dfrac{1}{n}x^2\right)^{\frac{n+1}{2}}}\cdot x$$

$$= \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\dfrac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

这正是 (5).

教学札记之二十七

$[0,1]$ 中数的 g 进制展开与概率论

在早期概率论的研究中, 发现有两个问题不能在离散样本空间中严格表述. 一个是几何概率中所称的从 $[0,1]$ 区间中随机选一个点, 另一个是可列重伯努利试验 E^∞, 其中样本空间都不可列. 例如庞加莱 (J. Poincaré, 1854 — 1912) 就在 1896 年出版的《概率演算》中提到在 $[0,1]$ 中随机选一点为有理数的概率. 后来博雷尔 (1871 — 1956) 在数学研究中涉及这两个问题, 发现它们是同一个问题, 于 1909 年发表了 "Les probabilités dénombrables et leurs applications arithmétiques".

在该文中, 博雷尔证明了 $[0,1]$ 中正规数 $\left(\text{即在该数的无尽十进制展开式中}, 0,1,2,\cdots,9\ \text{出现的频率均为}\ \dfrac{1}{10}\right)$ 的度量为 1, 这就是有名的强大数定律. 博雷尔的工作是第一次提出了以概率 1 收敛性.

博雷尔的研究揭示了概率论与测度论的内在的深刻联系, 极大地推动了科尔莫戈罗夫在博雷尔所创的测度论和他的学生勒贝格所创的测度 - 积分论的基础上建立起概率论的公理化体系.

一、二进制展开与博雷尔正规数定理

$[0,1]$ 中的任一个数 ω 可以用二进制展开成

$$\omega = 0\,.\,\xi_1(\omega)\xi_2(\omega)\xi_3(\omega)\cdots = \sum_{k=1}^{\infty}\frac{\xi_k(\omega)}{2^k} \tag{1}$$

其中 $\xi_k(\omega)$ 为 0 或 1.

有限小数如 0.1 有 $0.1000\cdots$ 和 $0.0111\cdots$ 两种无尽小数表示, 为使表示法唯一, 以后一律约定用后者, 余类推. 因此 $[0,1]$ 中的任一个数 ω 与无限序列 $(\xi_1(\omega),\xi_2(\omega),\xi_3(\omega),\cdots)$ 建立了一一对

应, 而后者可以表示一串掷硬币的可列重伯努利试验.

$[0,1]$ 原有的测度是长度, 对任何 $0<a<b<1$ 定义 $[a,b]$ 的长度为 $b-a$, 记作

$$P\{[a,b)\}=b-a \tag{2}$$

一般地, 对 $0 \leqslant a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n \leqslant 1$, 记 $A=\sum_{i=1}^{n}[a_i,b_i)$, 则

$$P\{A\}=\sum_{i=1}^{n}(b_i-a_i) \tag{3}$$

这个定义可以推广到 (严格说要用到测度扩张定理) 所有博雷尔点集, 甚至勒贝格点集 A, 仍记为 $P(A)$.

若记 $\Omega=[0,1]$, 记 $[0,1]$ 中一切博雷尔点集为 $\mathscr{F}$, 则 $(\Omega,\mathscr{F},P)$ 构成一个概率空间. 这个概率空间事实上在《概率论基础》的定理 3.3.1 的证明中, 例 5.3.4 及例 5.4.1 中已出现过. 由习题三 10 题知若是定义 $\eta(\omega)=\omega$, 则 $F(x)=P\{\eta(\omega)<x\}=x$ 为 $[0,1]$ 上均匀分布, 它给几何概率以严格的概率意义.

在展开式 (1) 中

$$\begin{aligned}
P\{\xi_1(\omega)=0\}&=P\left\{\omega\in\left(0,\frac{1}{2}\right]\right\}=\frac{1}{2}\\
P\{\xi_1(\omega)=1\}&=P\left\{\omega\in\left(\frac{1}{2},1\right]\right\}=\frac{1}{2}
\end{aligned} \tag{4}$$

类似地, 若 $a_1,a_2,\cdots,a_n$ 为任何一个 0-1 序列, 使 ω 满足

$$\sum_{i=1}^{n}\frac{a_i}{2^i}<\omega\leqslant\sum_{i=1}^{n}\frac{a_i}{2^i}+\sum_{i=n+1}^{\infty}\frac{1}{2^i}$$

则

$$\{\omega:\xi_i(\omega)=a_i,\ i=1,2,\cdots,n\}=\Big(\sum_{i=1}^{n}\frac{a_i}{2^i},\quad\sum_{i=1}^{n}a_i+\frac{1}{2^n}\Big]$$

从而

$$P\{\omega:\xi_i(\omega)=a_i,\ i=1,2,\cdots,n\}=\frac{1}{2^n} \tag{5}$$

特别地, $P\{\xi_i(\omega)=a_i\}=\frac{1}{2},\quad i=1,2,\cdots,n$, 故有

$$P\{\omega:\xi_i(\omega)=a_i,\quad i=1,2,\cdots,n\}=\prod_{i=1}^{n}P\{\omega:\xi_i(\omega)=a_i\} \tag{6}$$

这表明 $\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega)\cdots$ 相互独立, 各以概率 $\frac{1}{2}$ 取值 0 或 1, 可以看作掷一枚均匀硬币所对应的伯努利试验序列 E^∞ (从有限到无限严格说要用到科尔莫戈罗夫相容性定理).

这样一来, $[0,1]$ 均匀分布与 $p=\frac{1}{2}$ 的 E^∞ 被统一处理.

这时从《概率论基础》第二章知

$$P\left\{\omega:\sum_{i=1}^{n}\xi_i(\omega)=k\right\}=\binom{n}{k}\left(\frac{1}{2}\right)^n,\quad k=0,1,2,\cdots,n \tag{7}$$

从第五章 (5.1.14) 又知成立伯努利大数定律.

定理 1

$$\lim_{n\to\infty}P\left\{\omega:\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i(\omega)-\frac{1}{2}\right|\geqslant\varepsilon\right\}=0 \tag{8}$$

从第五章 (5.4.19) 知成立博雷尔强大数定律

定理 2

$$P\left\{\omega:\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\xi_i(\omega)=\frac{1}{2}\right\}=1 \tag{9}$$

这就是博雷尔当年证得的正规数定理, 也即 $p=\frac{1}{2}$ 的定理 5.4.2. 它表明在 $[0,1]$ 中的数的二进制展开式中 0, 1 出现各半的数 (正规数) 的测度为 1. 因为有限小数, 有理数全体都不是正规数, 其测度为 0, 这也回答了庞加莱的问题.

这是算术性的定理, 也是概率性的定理, 从现代角度看还是测度论性的定理. 这样一来, 博雷尔大大突破了早期概率论的研究范

围, 为概率论的发展开拓了新境界.

由于 $\left|\frac{\xi_i(\omega)}{2^i}\right| \leqslant \frac{1}{2^i}, \sum_{i=1}^{\infty}\frac{1}{2^i} < \infty$, 用魏尔斯特拉斯比较判别法知对每一个 $\omega \in \Omega, \sum_{i=1}^{\infty}\frac{\xi_i(\omega)}{2^i}$ 收敛, 故有

定理 3

$$P\left\{\lim_{n\to\infty}\sum_{i=1}^{n}\frac{\xi_i(\omega)}{2^i} = \omega\right\} = 1 \tag{10}$$

再由习题五 34 题可得

定理 4

$$\lim_{n\to\infty} P\left\{\sum_{i=1}^{n}\frac{\xi_i(\omega)}{2^i} < x\right\} = x, \quad 0 \leqslant x \leqslant 1 \tag{11}$$

即 $\sum_{i=1}^{n}\frac{\xi_i}{2^i}$ 的分布收敛于 $U[0,1]$, 其证明见习题解答, 用特征函数作工具, 关键是利用了下列韦达三角恒等式

$$\frac{\sin t}{t} = \prod_{k=1}^{\infty}\cos\frac{t}{2^k} \tag{12}$$

二、康托尔集与奇异函数

康托尔集的构造如下; 从区间 $[0,1]$ 中刨去三等分的中间开区间 $\left(\frac{1}{3},\frac{2}{3}\right)$, 剩下的是两个闭区间 $\left[0,\frac{1}{3}\right]$ 与 $\left[\frac{2}{3},1\right]$ 之并, 记之为 G_1, 为闭集. 接着, 对各闭区间三等分, 刨去中间的开区间 $\left(\frac{1}{3^2},\frac{2}{3^2}\right)$ 及 $\left(\frac{2}{3}+\frac{1}{3^2},\frac{2}{3}+\frac{2}{3^2}\right)$, 剩下的是 4 个互不相交的闭区间之并, 记之为 G_2, 也为闭集, 如图 5–3 所示.

如此进行下去, 可以得到单调下降的闭集序列 $\{G_n\}$. 记 $G = \bigcap_{n=1}^{\infty} G_n$, 则 G 是非空的闭集. 这个集称为**康托尔集**. 从 $[0,1]$ 刨去

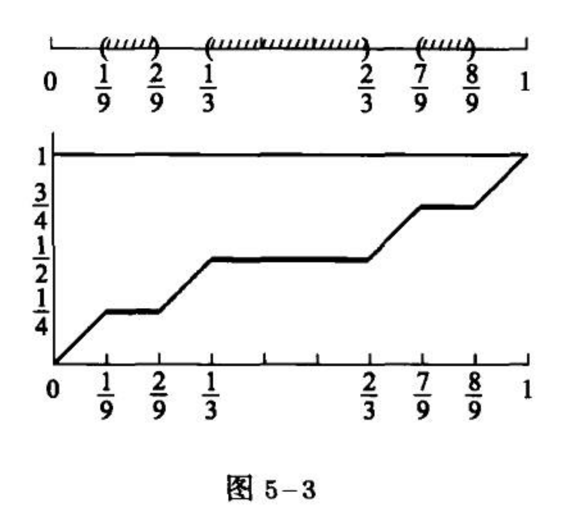


图 5-3

的集合 $\overline{G}$ 是可列个开区间之并, 是一维博雷尔点集, 从而康托尔集 G 也是一维博雷尔点集.

上述操作中, 第 n 阶段刨去 2^{n-1} 个开区间, 每个长度为 $\dfrac{1}{3^n}$, 刨去的总长度为

$$\sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 1 \tag{13}$$

因此康托尔集 G 的测度 (长度) 为 0.

对 $[0,1]$ 中的实数用三进制展开, 按康托尔集 G 的构成法, 第一阶段刨去开区间 $(0.1, 0.2)$, 第二阶段刨去开区间 $(0.01,\ 0.02)$ 及 $(0.21,\ 0.22), \cdots$, 第 n 阶段刨去的是形如

$$\left(0\,.\,a_1a_2\cdots a_{n-1}\,1,\quad 0\,.\,a_1a_2\cdots a_{n-1}\,2\right)$$

的开区间, 其中 a_i 取 0 或 2 的值全体. 因 G 中的点包含刨去的区间的端点, 这些端点在三进制表示中有两种表示法. 若约定

$$0\,.\,a_1a_2\cdots a_{n-1}\,1 = 0\,.\,a_1a_2\cdots a_{n-1}\,022\cdots$$

则这些端点也有唯一的表示法.

这样一来, G 中的数是 a_i 取 0 或 2 的三进制小数

$$0\,.\,a_1a_2a_3\cdots$$

反之, 能这样表示的实数也属于 G.

不要以为 G 中的数是一些孤立的点, 事实上

$$\sum_{k=1}^{\infty} \frac{a_k}{3^k} \longleftrightarrow \sum_{k=1}^{\infty} \frac{b_k}{2^k} \tag{14}$$

其中

$$b_k = \frac{a_k}{2} = \begin{cases} 0, & a_k = 0 \\ 1, & a_k = 2 \end{cases} \tag{15}$$

(14) 建立了 G 中的数与 $[0,1]$ 中数的一一对应, 因此 G 的势也是连续统, 而非可列点集.

在康托尔集 G 上定义函数 $F(x)$:

$$F(0\,.\,a_1a_2\cdots) = 0\,.\,b_1b_2\cdots \tag{16}$$

这里 $F(x)$ 之值以二进制表示. 这样一来,

$$\begin{aligned} F(0\,.\,a_1a_2\cdots a_{n-1}\,1) &= F(0\,.\,a_1a_2\cdots a_{n-1}\,022\cdots) \\ &= 0\,.\,b_1b_2\cdots b_{n-1}\,011\cdots \\ &= 0\,.\,b_1b_2\cdots b_{n-1}\,1 \\ &= F(0\,.\,a_1a_2\cdots a_{n-1}\,2) \end{aligned} \tag{17}$$

即 $F(x)$ 在刨去的区间的端点的值是相等的. 若以端点之值来定义 $F(x)$ 在这个刨去的区间中的函数值, 则这样定义的函数 $F(x)$ 是区间 $[0,1]$ 上单调增加的连续函数, 称为**康托尔函数**, 参见上文的图 5–3.

进一步, 若补充定义

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ 1, & x > 1 \end{cases}$$

则 $F(x)$ 为分布函数.

由于在 $\overline{G}$ 上, $F'(x) = 0$, 因此 $F(x)$ 只在康托尔集 G 上增加, 而康托尔集 G 的测度为 0, 因此这样定义的分布函数 $F(x)$ 是奇异函数.

在函数论中奇异函数是相当罕见的，所以在函数论或测度论中要举奇异函数的例子，几乎没有例外地都以康托尔函数为例. 但在概率论中两个测度 (或两个分布) 相互奇异是家常便饭，不过已超出我们讨论的范围.

三、二个奇异分布的卷积

两个分布函数的卷积对应于两个独立随机变量之和. 人们不禁要问: 两个奇异型独立随机变量之和属于何型?

当然可以是奇异型的，例如康托尔函数的卷积仍为奇异函数.

出人意料的是居然也可以是连续型. 下面略作介绍.

类似上段讨论，可以构造其他进制的康托尔分布. 例如

$$Z=\sum_{k=1}^{\infty}\frac{X_k}{4^k},\qquad X_k=\begin{cases}0, & \dfrac{1}{2}\\ 3, & \dfrac{1}{2}\end{cases} \tag{18}$$

是一个随机变量，它对应于把 $[0,1]$ 中的数作 4 进制展开，其 X_k 只包含 0 或 3 而无 1 或 2 的. 相当于通过把区间 $[0,1]$ 四等分，然后再刨去中间二段，依次不断进行得到的康托尔集.

另外，不难验证 $\begin{pmatrix}-1 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$ 的特征函数为 $\cos t$，而 $[-1,1]$ 的特征函数为 $\dfrac{\sin t}{t}$. 若 $\eta_k,\ k=1,2,\cdots$ 相互独立，均服从 $\begin{pmatrix}-1 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$，则 $U=\sum\limits_{k=1}^{\infty}\dfrac{\eta_k}{2^k}\sim U[-1,1]$.

注意这里用对称分布 $\begin{pmatrix}-1 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$ 代替前节 $\begin{pmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$ 以使特征函数为实值，便于处理，事实上这里用的是伯努利试验的另一种表达法，在随机游动中曾用过. 这里的 $\eta_k=2\xi_k-1$，又称为

Rademacher 随机变量. 两种表达法的等价性是明显的.

因此韦达公式 $\dfrac{\sin t}{t} = \prod_{k=1}^{\infty} \cos \dfrac{t}{2^k}$ 是 $U = \sum_{k=1}^{\infty} \dfrac{\eta_k}{2^k}$ 的特征函数表示.

显然

$$\prod_{k=1}^{\infty} \cos \frac{t}{2^k} = \prod_{k=1}^{\infty} \cos \left(2 \cdot \frac{t}{4^k}\right) \prod_{k=1}^{\infty} \cos \left(\frac{t}{4^k}\right) \tag{19}$$

$$U \overset{\text{def}}{=\!=} 2V_1 + V_2 \tag{20}$$

右边两个无穷乘积分别表示两个奇异型随机变量的特征函数, 这两个独立随机变量之和是连续型的, 因为左边表示 $U[-1,1]$ 的特征函数.

同样地

$$\prod_{k=1}^{\infty} \cos \frac{t}{3^k} = \prod_{k=1}^{\infty} \cos \left(3 \cdot \frac{t}{9^k}\right) \prod_{k=1}^{\infty} \cos \frac{t}{9^k} \tag{21}$$

$$Y \overset{\text{def}}{=\!=} 3W_1 + W_2$$

若以 C, D, S 分别记连续型, 离散型, 奇异型, 则独立随机变量 X 与 Y 之和的型如下:

X \ Y	C	D	S
C	C	C	C
D	C	D	S
S	C	S	C 或 S

教学札记之二十八

切比雪夫不等式与大数定律的证明

1866 年切比雪夫发表了论文 "论平均值". 这是一篇开创性的论文. 从现代的观点来看, 这篇论文证明了一个不等式——**切比雪**

夫不等式; 用这个不等式证明了一条大数定律——**切比雪夫大数定律**, 它只要求方差有界, 把伯努利大数定律和泊松大数定律都作为一个很普通的特例. 同样重要的是这篇论文使用随机变量的概念, 把数字特征作为工具, 是"矩法"的开山之作. 从此, 随机变量成了表述概率问题的主要工具, 古典型的框框被彻底突破, 概率论的发展中心也从西欧转移到东欧的圣彼得堡.

新方法展现出巨大的威力, 1887 年切比雪夫又开创了方差有界场合一般中心极限定理的研究, 虽然结果不够完善, 但指明了方向. 下一步由他的两位杰出的学生来完成: 马尔可夫 (1856 — 1922) 拓展矩法, 研究了相依变量的极限定理, 开创了马尔可夫过程这一崭新领域; 而李雅普诺夫 (1857 — 1918) 则引进特征函数这一有力工具, 证得了李雅普诺夫中心极限定理, 为古典极限定理的最终解决奠下基石. 他们的继承者辛钦和科尔莫戈罗夫在近代极限定理的研究中更是名垂青史.

一、切比雪夫不等式

1866 年切比雪夫建立了不等式

$$P\{|\xi - E\xi| \geqslant \varepsilon\} \leqslant \frac{D\xi}{\varepsilon^2} \tag{1}$$

他首先就把它用来研究大数定律. 为节省篇幅, 本文一律沿用《概率论基础》第五章的符号, 不再一一说明.

当时已知的大数定律, 一个是众口称誉的建立于 150 多年前的**伯努利大数定律**:

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right\} = 1 \tag{2}$$

另一个是不久前才提出, 当时很少有人相信的**泊松大数定律**:

$$\lim_{n\to\infty} P\left\{\left|\frac{\mu_n}{n} - \frac{p_1 + p_2 + \cdots + p_n}{n}\right| < \varepsilon\right\} = 1 \tag{3}$$

这两个大数定律都是使用复杂的分析方法证明的, 若使用切

比雪夫不等式, 它们的证明只要一行文字:

$$P\left\{\left|\frac{\mu_n}{n}-\frac{p_1+p_2+\cdots+p_n}{n}\right|\geqslant\varepsilon\right\}\leqslant\frac{D\left(\frac{\mu_n}{n}\right)}{\varepsilon^2}=\frac{D(\mu_n)}{n^2\varepsilon^2}$$

$$=\frac{\sum\limits_{k=1}^{n}p_kq_k}{n^2\varepsilon^2}\leqslant\frac{1}{4n\varepsilon^2}$$

切比雪夫注意到加项 0-1 分布的假设并非必要, 独立性也可放宽, 于是建立自己的大数定律:

若 $\{\xi_k,\ k=1,2,\cdots\}$ 两两不相关, 且 $D\xi_k\leqslant C,\ k=1,2,\cdots$, 则

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|<\varepsilon\right\}=1 \tag{4}$$

按现代的记号, 证明也只要一行:

$$\begin{aligned}&P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k\right|\geqslant\varepsilon\right\}\\&\leqslant\frac{1}{\varepsilon^2}D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n}D\xi_k\leqslant\frac{C}{n\varepsilon^2}\longrightarrow 0\end{aligned} \tag{5}$$

随机变量的表示、切比雪夫不等式、矩法等新方法显示出无穷的威力.

首创者过于着眼要解决的具体问题, 后继者更关心新方法的适用范围. 马尔可夫详细地审视了老师的证明, 发现 (4) 式的成立, 不相关和方差一致有界的假定并非必要, 只要在 (5) 中直接要求

$$\frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right)\longrightarrow 0 \tag{6}$$

即可. 这就是著名的**马尔可夫条件**, 是大数定律成立的一个最常用的充分条件, 从而建立了自己的大数定律即**马尔可夫大数定律**.

舍去独立性, 舍去方差的一致有界性, 从而使大数定律的应用范围大为推广. 过渡到非独立场合特别有意义, 因近代 (例如随机

过程论中) 要研究有某种相依性的随机变量序列的极限定理, 马尔可夫在此领域成了首创者.

二、马尔可夫不等式

马尔可夫在这个过程中还创立了自己的不等式, 即**马尔可夫不等式**

$$P\{|\xi| \geqslant \varepsilon\} \leqslant \frac{E|\xi|^r}{\varepsilon^r}, \quad r > 0 \tag{7}$$

此即《概率论基础》中的 (5.2.22), 课文中的证明过程故意采用与切比雪夫不等式的证明过程表面相反但实质一致的顺序, 请读者细察.

当 $r = 2$ 即化为切比雪夫不等式. 但适当的 r 在某些场合是必要的.

下面介绍三个应用方面的例子.

1. 当把 2 阶矩推广到 $r\,(r > 0)$ 阶矩, 马尔可夫不等式一锤定音地证明了定理 5.2.8 即 r 阶收敛推出依概率收敛. 事实上

$$P\{|\xi_n - \xi| \geqslant \varepsilon\} \leqslant \frac{E|\xi_n - \xi|^r}{\varepsilon^r} \longrightarrow 0 \tag{8}$$

2. 在博雷尔强大数定律证明中的应用.

为证明定理 5.4.2 (**博雷尔强大数定律**)

$$P\left\{\frac{\mu_n}{n} \longrightarrow p\right\} = 1 \tag{9}$$

必须证明

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{\mu_n}{n} - p\right| \geqslant \varepsilon\right\} < \infty \tag{10}$$

这时若利用切比雪夫不等式, 只能证得

$$P\left\{\left|\frac{\mu_n}{n} - p\right| \geqslant \varepsilon\right\} \leqslant \frac{1}{4n\varepsilon^2} \tag{11}$$

这不足以保证 (10) 成立, 必须有更好的估计方法, 这时利用 $r = 4$

的马尔可夫不等式就能证得

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^4}E\left|\frac{\mu_n}{n}-p\right|^4<\frac{1}{\varepsilon^4 4n^2} \tag{12}$$

从而保证了 (10) 式的成立, 并证得定理.

3. 习题五 46 题要求读者把博雷尔强大数定律推广到独立同分布有有限 4 阶矩的场合, 证法类似.

事实上, 作更细致的考察与计算, 能推广到独立而非同分布但 4 阶矩有限的场合, 即著名的康特立强大数定律, 详细叙述如下:

康特立强大数定律: 设 $\xi_1,\xi_2,\cdots,\xi_n\cdots$ 是独立随机变量序列, $E|\xi_i-E\xi_i|^4\leqslant C$, 其中 C 为某个常数, $i=1,2,\cdots$, 则

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-E\xi_i)=0\right\}=1 \tag{13}$$

证明 为书写方便, 不失一般性, 设 $E\xi_i=0,\quad i=1,2,\cdots$, 记 $S_n=\xi_1+\cdots+\xi_n$,

$$\begin{aligned}S_n^4&=(\xi_1+\cdots+\xi_n)^4\\&=\sum_{i=1}^{n}\xi_i^4+\sum_{i<j}\frac{4!}{2!\,2!}\xi_i^2\xi_j^2+\sum_{\substack{i\neq j\\ i\neq k\\ j<k}}\frac{4!}{2!\,1!\,1!}\xi_i^2\xi_j\xi_k\\&\quad+\sum_{i<j<k<l}4!\,\xi_i\xi_j\xi_k\xi_l+\sum_{i\neq j}\frac{4!}{3!\,1!}\xi_i^3\xi_j\end{aligned}$$

故

$$\begin{aligned}ES_n^4&=\sum_{i=1}^{n}E\xi_i^4+6\sum_{i<j}E\xi_i^2\,E\xi_j^2\\&\leqslant nC+6\sum_{i<j}\sqrt{E\xi_i^4\,E\xi_j^4}\\&\leqslant nC+\frac{6n(n-1)}{2}C<3n^2C\end{aligned} \tag{14}$$

因此

$$\sum_{n=1}^{\infty} E\left(\frac{S_n}{n}\right)^4 \leqslant 3C\sum_{n=1}^{\infty}\frac{1}{n^2} < \infty \tag{15}$$

由马尔可夫不等式

$$P\left\{\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^4} E\left|\frac{S_n}{n}\right|^4$$

立即得到

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{S_n}{n}\right| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^4}\sum_{n=1}^{\infty} E\left|\frac{S_n}{n}\right|^4 < \infty \tag{16}$$

由博雷尔–康特立引理知

$$\frac{S_n}{n} \xrightarrow{a.s.} 0$$

这个证明中, 用切比雪夫不等式只能处理 2 阶矩是不够的, 用马尔可夫不等式估计 4 阶矩正好. 这个定理包含博雷尔强大数定律为特例.

但在更一般的场合, 为证明强大数定律成立, 须证明

$$\lim_{m\to\infty} P\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i - E\xi_i)\right| \geqslant \varepsilon\right\} = 0 \tag{17}$$

这时马尔可夫不等式就无能为力了. 这个难关由他们的后辈科尔莫戈罗夫突破.

三、科尔莫戈罗夫不等式

科尔莫戈罗夫不等式 设 $\xi_1, \xi_2, \cdots, \xi_n$ 是独立随机变量, 方差有限, 则对任意 $\varepsilon > 0$ 成立

$$P\left\{\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{j}(\xi_i - E\xi_i)\right| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^2}\sum_{j=1}^{n} D\xi_j \tag{18}$$

当 $n=1$ 时得到切比雪夫不等式, 所以它是切比雪夫不等式往另一方向的推广.

科尔莫戈罗夫不等式对随机变量部分和的最大值作概率估计，含意深刻，证明难度大，用途也广，在强大数定律的证明中起主要作用. 关于科尔莫戈罗夫不等式已有许许多多的推广，每出现一种随机过程差不多都要推导相应的不等式.

总之，从切比雪夫不等式开始的不等式林林总总不下数十种，要写一本关于这类概率不等式的书一定相当厚.

值得指出，切比雪夫类型不等式在概率论中的应用并不局限于大数定律. 另外关于科尔莫戈罗夫不等式在强大数定律中的应用这里也不展开了，部分内容已写入教学札记之二十九“强大数定律证明途径评述”.

这段历史的回顾让人印象深刻的是学派的创立，研究课题和方法的延续和创新.

四、一点注记

切比雪夫不等式可写成

$$P\left\{\frac{|\xi - E\xi|}{\sqrt{D\xi}} \geqslant \delta\right\} \leqslant \frac{1}{\delta^2} \tag{19}$$

对每个 $\delta > 0$ 成立.

由于 $\delta < 1$ 肯定成立，因此只要考虑 $\delta \geqslant 1$ 的情况.

能构造一个使等号成立的随机变量，例如

ξ	$-t$	0	t
P	$\frac{1}{2t^2}$	$1-\frac{1}{t^2}$	$\frac{1}{2t^2}$

$$E\xi = 0, \quad D\xi = 1, \quad P\{|\xi| \geqslant t\} = \frac{1}{t^2}$$

在这个意义下切比雪夫不等式已是最好的估计，无法再改进. 但这个随机变量概率分布随 t 而变，因此对于一个给定的随机变量它并不一定是最好的. 进一步，若已知分布，则 (19) 式左端的概率不需要任何估计. 以正态分布为例说明之，这时用切比雪夫不等式

估计的效果欠佳.

δ	1	2	3	4
所有分布	1.00	0.25	0.11	0.06
正态分布	0.313 3	0.045 5	0.002 7	0.000 06

概率论历史研究表明, 切比雪夫不等式和矩法先为法国数学家比埃奈梅 (Bienaymé, 1796—1878) 发现. 事实上,切比雪夫本人也把发明权归给他.

教学札记之二十九

强大数定律证明途径评述

对于独立同分布场合, 科尔莫戈罗夫证得数学期望的存在是强大数定律成立的充分必要条件, 见《概率论基础》定理 5.4.4. 这是一个最终的结果.

但对于一般独立随机变量序列则得不到如此漂亮的结果, 因此只好找一个优异而又易于验证的充分条件. 目前最好的结果也属于科尔莫戈罗夫, 这便是通常所讲的科尔莫戈罗夫强大数定律. 我们下面要评述的便是这个强大数定律的证明途径, 因为它有极大创意而且充满技巧性. 在《概率论基础》中它与最后一节的林德伯格-费勒中心极限定理的证明技巧是概率论中的证明方法的典范. 况且上述定理 5.4.4 也是利用这个定理加"截尾法"证明的.

科尔莫戈罗夫强大数定律 设 $\{\xi_i,\ i=1,2,\cdots\}$ 是独立随机变量序列, 若

$$\sum_{n=1}^{\infty}\frac{D\xi_n}{n^2}<\infty \tag{1}$$

则成立

$$P\Big\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-E\xi_i)=0\Big\}=1 \tag{2}$$

对于 (2) 这种以概率 1 收敛性, 课文 (5.4.16) 已指明相当于要求证明

$$\lim_{k\to\infty} P\Big\{\max_{1\leqslant n\leqslant k}\Big(\frac{1}{n}\sum_{i=1}^{n}|\xi_i - E\xi_i|\Big)\geqslant\varepsilon\Big\}=0 \tag{3}$$

为了用矩来表达上述收敛条件, 必须用 "矩法" 建立起切比雪夫型不等式. 但这里遇到的是事件 $A_n=\Big\{\frac{1}{n}\sum_{i=1}^{n}|\xi_i - E\xi_i|\geqslant\varepsilon\Big\}$ 的并 $\bigcup_{n=1}^{k}A_n$, 因此切比雪夫不等式是不够用的, 这时科尔莫戈罗夫建立了新型不等式, 现在已通称科尔莫戈罗夫不等式, 是切比雪夫型不等式的重大发展.

科尔莫戈罗夫不等式 设 $\xi_1,\xi_2,\cdots,\xi_n$ 是独立随机变量, 方差有限, 则对任意 $\varepsilon>0$, 成立

$$P\Big\{\max_{1\leqslant j\leqslant n}\Big|\sum_{i=1}^{j}(\xi_i - E\xi_i)\Big|\geqslant\varepsilon\Big\}\leqslant\frac{1}{\varepsilon^2}\sum_{j=1}^{n}D\xi_j \tag{4}$$

当 $n=1$ 时就得到切比雪夫不等式.

但是科尔莫戈罗夫不等式比切比雪夫不等式强很多. 以这里讨论的 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立场合为例. 切比雪夫不等式给出:

$$P\Big\{\Big|\sum_{i=1}^{n}(\xi_i - E\xi_i)\Big|\geqslant\varepsilon\Big\}\leqslant\frac{1}{\varepsilon^2}\sum_{i=1}^{n}D\xi_i \tag{5}$$

而科尔莫戈罗夫不等式则对更大的集合 $\bigcup_{j=1}^{n}\Big\{\Big|\sum_{i=1}^{j}(\xi_i - E\xi_i)\Big|\geqslant\varepsilon\Big\}$ 的概率给出同样的上界.

科尔莫戈罗夫不等式证明的关键在于把如下的复杂事件 $\Big\{\max_{1\leqslant j\leqslant n}\Big|\sum_{i=1}^{j}(\xi_i - E\xi_i)\Big|\geqslant\varepsilon\Big\}$ 用 "首次进入 (即 $\geqslant\varepsilon$) 分解法" (参看习题一 4 题) 表示成不相容事件之和, 再利用随机变量的独立性使用马尔可夫不等式过渡到 "矩" 而得到.

科尔莫戈罗夫不等式已被推广到各种场合, 出现许多变形.

总的说, 要证明强大数定律离不开科尔莫戈罗夫型不等式, 再配合上 (3) 这类收敛性条件 (博雷尔–康特立引理是作用相当的另一种形式), 还使用一点别的技巧就能达到目的.

目前常见的证法大体有三种, 简介如下:

1. 普通方式 (见于大多数教科书)

证明要点: 由科尔莫戈罗夫不等式知: 对任意 $a > 0$,

$$P\left\{\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{j}\frac{\xi_i - E\xi_i}{i}\right| \geqslant a\right\} \leqslant \frac{1}{a^2}\sum_{i=1}^{n}\frac{D\xi_i}{i^2} \tag{6}$$

因此

$$P\left\{\max_{j\geqslant 1}\left|\sum_{i=1}^{j}\frac{\xi_i - E\xi_i}{i}\right| \geqslant a\right\} \leqslant \frac{1}{a^2}\sum_{i=1}^{\infty}\frac{D\xi_i}{i^2} \tag{7}$$

故

$$\begin{aligned} P\left\{\sum_{i=1}^{\infty}\frac{\xi_i - E\xi_i}{i} < a\right\} &\geqslant P\left\{\max_{j\geqslant 1}\left|\sum_{i=1}^{j}\frac{\xi_i - E\xi_i}{i}\right| < a\right\} \\ &\geqslant 1 - \frac{1}{a^2}\sum_{i=1}^{\infty}\frac{D\xi_i}{i^2} \end{aligned} \tag{8}$$

因 $\sum\limits_{i=1}^{\infty}\frac{D\xi_i}{i^2} < \infty$, 令 $a \longrightarrow \infty$ 得

$$P\left\{\sum_{i=1}^{\infty}\frac{\xi_i - E\xi_i}{i} < \infty\right\} = 1 \tag{9}$$

这时用一条数学分析中有名的 **Kronecker 引理**:

若 $\sum\limits_{k=1}^{\infty}\frac{c_k}{k} < \infty$, 则 $\frac{1}{n}\sum\limits_{k=1}^{n}c_k \longrightarrow 0$,

即得

$$\frac{1}{n}\sum_{i=1}^{n}(\xi_i - E\xi_i) \xrightarrow{a.s.} 0 \tag{10}$$

利用科尔莫戈罗夫不等式证明条件 (1) 可保证 $\sum_{i=1}^{\infty} \frac{\xi_i - E\xi_i}{i}$ 以概率 1 收敛, 再用 Kronecker 引理推知 $\frac{1}{n}\sum_{i=1}^{n}(\xi_i - E\xi_i) \xrightarrow{a.s.} 0$, 是这种方法的主要证明思路.

2. 用推广的博雷尔 – 康特立引理

对 $\{A_n\}$ 找一个整数序列 $1 = n_1 < n_2 < \cdots$, 注意到

$$\sum_{j=1}^{\infty} P\Big\{\bigcup_{n=n_j}^{n_{j+1}-1} A_n\Big\} \leqslant \sum_{j=1}^{\infty}\sum_{n=n_j}^{n_{j+1}-1} P(A_n) = \sum_{n=1}^{\infty} P(A_n)$$

可看出 $\sum_{j=1}^{\infty} P\Big\{\bigcup_{n=n_j}^{n_{j+1}-1} A_n\Big\} < \infty$ 比 $\sum_{n=1}^{\infty} P(A_n) < \infty$ 稍弱.

推广的博雷尔 – 康特立引理: 若 $\sum_{j=1}^{\infty} P\Big\{\bigcup_{n=n_j}^{n_{j+1}-1} A_n\Big\} < \infty$, 则 $P\{\overline{\lim_{n\to\infty}} A_n\} = 0$ (证明略).

利用科尔莫戈罗夫不等式得

$$\begin{aligned} p_m &= P\Big\{\max_{2^m \leqslant n < 2^{m+1}} \Big(\frac{1}{n}\Big|\sum_{k=1}^{n}(\xi_k - E\xi_k)\Big|\Big) \geqslant \varepsilon\Big\} \\ &\leqslant \frac{1}{\varepsilon^2 2^{2m}} \sum_{j<2^{m+1}} D\xi_j \end{aligned} \tag{11}$$

而

$$\begin{aligned} \sum_{m=1}^{\infty} p_m &\leqslant \sum_{m=1}^{\infty} \frac{1}{\varepsilon^2 2^{2m}} \sum_{j<2^{m+1}} D\xi_j \\ &= \frac{1}{\varepsilon^2}\sum_{j=1}^{\infty} D\xi_j \sum\nolimits_m \frac{1}{2^{2m}} \end{aligned} \tag{12}$$

这里 $\sum_m$ 对 $2^{m+1} > j$ 的 m 值求和.

以不等式 $2^p \leqslant j < 2^{p+1}$ 来定 p, 于是

$$\sum\nolimits_m 2^{-2m} = \sum_{m=p}^{\infty} 2^{-2m} = \frac{4}{3} \cdot 2^{-2p} < \frac{16}{3j^2}$$

因此当 $\sum_{j=1}^{\infty} \frac{D\xi_j}{j^2} < \infty$ 时,

$$\sum_{m=1}^{\infty} p_m \leqslant \frac{16}{3\varepsilon^2} \sum_{j=1}^{\infty} \frac{D\xi_j}{j^2} < +\infty \tag{13}$$

用推广的博雷尔 – 康特立引理知

$$\frac{1}{n} \sum_{k=1}^{n} (\xi_k - E\xi_k) \geqslant \varepsilon \tag{14}$$

发生无限次的概率为 0, 即以概率 1 成立

$$\frac{1}{n} \sum_{k=1}^{n} (\xi_k - E\xi_k) \longrightarrow 0$$

从而证得强大数定律.

这种证法初见于格涅坚科《概率论教程》, 后为国内一些教材所采用.

3. 通过推广科尔莫戈罗夫不等式

《概率论基础》所采用的证法就属于这一类. 这时先把科尔莫戈罗夫不等式推广为**噶依克 – 瑞尼不等式**

$$P\Big\{ \max_{m \leqslant j \leqslant n} C_j \Big| \sum_{i=1}^{j} (\xi_i - E\xi_i) \Big| \geqslant \varepsilon \Big\} \leqslant \frac{1}{\varepsilon^2} \Big(C_m^2 \sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{n} C_j^2 \sigma_j^2 \Big)$$

为证明这个不等式可先把事件 $\Big\{ \max_{m \leqslant j \leqslant n} C_j \Big| \sum_{i=1}^{j} (\xi_i - E\xi_i) \Big| \geqslant \varepsilon \Big\}$ 用 "首次进入分解法" 表示成不相容事件之和, 再巧用随机变量的独立性, 利用马尔可夫不等式过渡到 "矩" 而得到 (见课文).

当 $m = 1$, $C_j = 1$ 即为科尔莫戈罗夫不等式.

取 $C_j = \frac{1}{j}$, 利用这个不等式比较快就证得科尔莫戈罗夫强大数定律.

《概率论基础》编写时从 Rao 和 Rohatgi 的书中取来这种比较新颖的证法, 而把另外两种证法改编成五道加双星号的习题. 这次出第三版, 因考虑这些题目难度过高, 已经删去.

科尔莫戈罗夫强大数定律还有其他证法, 而且可预计还会有新的证法不断推出. 希望这个评述能帮助读者对这个定理的证明加深理解.

教学札记之三十

关于进一步学习的建议

学过《概率论基础》之后, 许多读者会有进一步学习的要求, 或许希望听取一些建议. 因此在这里谈些个人的看法. 但本文受笔者的学识所限, 所提建议未必正确, 谨供参考.

我想应把读者区分为两类. 一类是准备以概率、统计为专业或从事与概率、统计关系极其密切的专业的读者; 另一类是准备以概率、统计为工具从事自己专业工作的.

对第二类读者, 我认为如果他们已掌握了《概率论基础》的基本内容, 那么在概率论知识方面基本上是够用的. 为开展在本学科的研究, 应该学习数理统计, 向应用方面发展. 这方面内容很多, 够学很长一段时间. 如果碰到有概率论方面的知识不足, 那时再回头来补习, 时犹未晚, 也更有针对性.

在学完数理统计的公共部分之后, 还要按照专业方向学习一些专题即数理统计的某些分支. 大体讲, 对从事经济管理研究的, 回归分析和时间序列分析最为重要. 对从事社会学研究的, 则抽样调查和相关性指标 (属非参数统计) 比较有用. 对从事工业技术的可关心质量管理. 对学医学的涉及统计知识很多, 例如应多关心列联表、广义线性模型和生存分析, 有些场合, 多元分析是有力工具.

总的说对这类读者，透彻理解统计学思想，掌握某些分支的知识，熟练掌握一两种软件系统就可以在本专业做出好的成绩.

对于第一类读者，应当学习测度论，照我的看法也应当学习数理统计，完全不懂统计，很难真正理解概率论，少数人除外.

这里所讲的测度论，事实上是指测度论与概率论结合起来讲的以测度论为基础的高等概率论，名字不少就叫 "高等概率论" 或 "现代概率论" (事实上目前学测度论的人绝大部分是学概率统计的学生). 这种书国内出版的至少有十本，国外好的至少也有十本，可选择一本学习.

学完测度论的高等概率论之后就可以进入到概率论的各个领域，主要是随机过程、随机分析、极限定理和各应用分支.

学习数理统计，绝大部分书籍并不假定要懂测度论，但进入数理统计比较理论或高深的部分还是非要测度论知识不可. 邵军 (Jun Shao) 的新作 "Mathematical Statistics" (Springer)，可为首选.

顺便提及，若要进入蒙特卡罗，刘军 (Jun S. Liu)的书 "Monte Carlo Strategies in Scientific Computing" 可为首选.

学习极限定理，古典的还是要读格涅坚科和科尔莫戈罗夫的书，有中译本，但钟开莱的英译本加了很好的注释可参考. 比较现代的内容可看佩特罗夫的专著，也有中译本《独立随机变量之和的极限定理》.

随机过程的入门书，Ross 的书 "Stochastic Processes"有特色，其次是 Karlin & Taylor 的几本书.

随机过程的主体是马尔可夫过程，从离散时间马尔可夫链开始到连续时间马尔可夫链，钟开莱的书是经典，要进入排队论也应从这里开始.

连续时间马尔可夫过程，Doob 的书是经典，Feller 第二卷也有不少内容. 但还是选一本较新的入手，例如陈木法的书.

近年来随机分析受到最大重视，书出得最多，鞅、布朗运动、随机积分、随机微分方程都列入此类. 想进入数理金融学的要有相当

的随机分析的基础, 不过这几年已出了近十本直接把数理金融学与随机分析合在一起讲的书, 可选其中一本进入此一领域.

从应用上讲, 平稳过程结合时间序列分析从通信领域到经济领域都有应用. Brockwell and Davis “Introduction to Time Series and Forecasting” 有一定理论深度. 当然也可以找应用性的书入门.

概率论发展历史年表

随机性问题起源很早: 近似于骰子的玩具在公元前 5000 年的遗址中出土; 人口调查在公元前 2700 年已进行; 14 世纪意大利和荷兰分别成立第一家航海保险公司; 关于偶然性与必然性的关系在古代就是哲学课题, 但近代意义的概率论起源于 1654 年.

1494 年 意大利数学家帕乔利出版《算术、几何、比与比例集成》, 内有"点数问题".

1539 年 卡尔达诺讨论"点数问题".《论机会游戏》约写于 1520 年但直至 1663 年才出版.

1556 年 塔塔利亚出版《论数学与测量》, 讨论"点数问题".

1642 年 伽利略 (1564—1642) 写出论文"掷骰子点数的思考" (1718 年它在佛罗伦萨第一次出版, 写作年代不确知), 论述为何掷三颗骰子出 10 点比出 9 点次数多.

1654 年 帕斯卡与费马通信, 两人用不同方法都正确解答了"点数问题", 标志概率论诞生.

1657 年 荷兰惠更斯出版《论赌博中的计算》, 引进数学期望概念, 提出"赌徒输光问题".

1662 年 英国格朗特发表《根据死亡公告作的自然和政治的观察》, 开创统计学新时代.

1665 年 牛顿提出第一个几何概率问题.

1669 年 惠更斯根据格朗特著作构造了一根死亡率曲线, 开创应用概率论于人口统计.

1693 年 英国哈雷发表"从布雷斯劳市的出生、葬礼表中估计人类的死亡率; 尝试确定年金价格".

1705 年 雅科布·伯努利去世, 其著作《猜度术》于 1713 年由其子侄整理出版.

1708 年 法国数学家蒙特莫特出版《随笔》, 1713 年出第二版, 给出"点数问题"两种一般表达式.

1711 年 棣莫弗发表"关于游戏中机遇巧合的概率".

1718 年 棣莫弗出版《机会学说》.

1726 年 尼古拉·伯努利提出"圣彼得堡悖论".

1733 年 棣莫弗导出正态曲线, 提出最初的"中心极限定理".

1738 年 丹尼尔·伯努利发表"风险测度的新理论注释".

1755 年 英国数学家辛普森发表"在应用天文学中取若干个观察值的平均的好处". 他在概率论中第一个引进连续型分布.

1761 年 贝叶斯去世, 其遗作"论机会学说中的一个问题"于 1763 年发表.

1770 年代 欧拉对彩票、人口统计和保险方面做了许多研究.

1777 年 蒲丰提出"投针问题".

1783 年 拉普拉斯建议用正态曲线表示误差分布概率.

1802 年 拉普拉斯主持法国人口选样调查.

1805 年 勒让德发表论文"确定彗星轨道的新方法", 提出最小二乘法.

1809 年 高斯出版《天体运动理论》, 导出正态分布.

1812 年 拉普拉斯出版《概率论的分析理论》.

1821 年 高斯出版《观察误差的组合理论》.

1827 年 英国植物学家布朗发现"布朗运动".

1829 年 法国比埃奈梅开始发表社会学和人口统计学论文.

1835 年 比利时凯特勒出版《论人类》, 提出"平均人"概念.

1837 年 泊松出版《关于犯罪和民事判决的概率研究》.

1851 年 第一届国际统计会议在布鲁塞尔召开.

1852 年 比埃奈梅导出 χ^2 分布.

1853 年 柯西在研究最小二乘法反例时提出柯西分布.

1859 年 麦克斯韦提出气体分子速度分布.

1859 年 达尔文出版《物种起源》.

1865 年 孟德尔发表《植物杂交试验》, 提出遗传学两条定律.

1866 年 切比雪夫发表"论平均值".

1869 年 高尔顿出版《遗传基因》, 开创生物统计研究, 引进相关和回归, 创用中位数, 百分位数.

1873 年 高尔顿提出分支过程和灭种概率.

1885 年 国际统计学会成立.

1889 年 高尔顿出版《自然遗传》.

1893 年 皮尔逊第一次使用术语"矩", 引进"标准差"术语和记号 σ, 又在一次演讲中命名高斯曲线为"正态分布".

1896 年 贝克勒尔发现放射性衰变现象.

1898 年 马尔可夫用矩法证明中心极限定理.

1899 年 贝特朗提出"贝特朗奇论".

1900 年 法国巴舍利耶发表《投机理论》, 用布朗运动描述股票价格.

1900 年 李雅普诺夫用特征函数法证明"李雅普诺夫定理".

1900 年 皮尔逊提出 χ^2 检验法, 开创近代抽样理论研究.

1900 年 希尔伯特第六问题提出物理学科 (力学与概率论) 的公理化.

1900 年 普朗克提出量子论.

1901 年 英国 Biometrika 创刊.

1905 年 爱因斯坦对布朗运动作出定性解释和定量描述.

1906 年 马尔可夫提出马尔可夫链.

1908 年 哥塞特发表 t 分布.

1909 年 埃尔朗研究电话交换模型.

1909 年 博雷尔提出强大数定律.

1910 年 英国尤尔出版《统计学导论》, 首次区分"总体"和"样本".

1915 年起 费希尔导出样本相关系数的抽样分布, 之后领导统计

学发展整整 20 年.

1918 年 维纳提出布朗运动的数学模型.

1918 年 费希尔引进"方差"这一名称.

1922 年 费希尔出版《理论统计学的数学基础》.

1922 年 林德伯格提出"林德伯格条件".

1923 年 费希尔提出""方差分析法".

1923 — 1927 年 量子力学建立.

1925 年 海森伯提出矩阵力学.

1926 年 薛定谔提出波动方程.

1926 年 玻恩提出概率诠释.

1927 年 海森伯提出测不准关系式.

1925 年 费希尔出版《供研究人员用的统计方法》.

1927 年 英国 Tippert 发表随机数表.

1928 年 冯·米泽斯出版《概率、统计和真理》.

1929 年 科尔莫戈罗夫发表"一般测度论和概率论".

1930 年 费希尔出版《自然选择的遗传理论》.

1931 年 科尔莫戈罗夫发表"概率论中的解析方法".

1931 年 国际"计量经济学会"成立.

1933 年 Econometrica 创刊.

1933 年 科尔莫戈罗夫出版《概率论基础》.

1934 年 莱维给出无穷可分分布律的刻画.

1934 年 辛钦引入平稳随机过程.

1935 年 费希尔出版《试验设计》.

1937 年 莱维出版《随机变量之和的理论》.

1942 年 科尔莫戈罗夫发表"希尔伯特空间的平稳序列".

1942 年 日本伊藤清提出 Itô 积分.

1945 年 美国造出世界第一台电子计算机 ENIAC.

1948 年 莱维出版《随机过程与布朗运动》.

1948 年 香农创立信息论.

1949 年 维纳的著作《平稳序列的外推, 内插和平滑》解密发表.

1949 年 格涅坚科与科尔莫戈罗夫出版《相互独立随机变数之和的极限分布》.

1950 年 费勒《概率论及其应用》第一卷出版.

1953 年 杜布出版《随机过程论》.

参 考 书 目

[1] 李贤平. 概率论基础. 第三版. 北京: 高等教育出版社, 2010.

[2] 李贤平, 沈崇圣, 陈子毅. 概率论与数理统计. 上海: 复旦大学出版社, 2003.

[3] 格涅坚科. 概率论教程. 丁寿田, 译. 北京: 高等教育出版社, 1956.

[4] 施利亚耶夫. 概率. 第一卷, 第二卷. 周概容等, 译. 北京: 高等教育出版社, 2007.

[5] 费勒. 概率论及其应用. 第一卷. 胡迪鹤, 林向清, 译. 第二卷. 刘文, 译. 北京: 科学出版社, 1964, 1994.

[6] 王梓坤. 概率论基础及其应用. 北京: 科学出版社, 1976.

[7] 钟开莱. 初等概率论附随机过程. 魏宗舒等, 译. 北京: 人民教育出版社, 1979.

[8] 李贤平, 卞国瑞, 吴立鹏. 概率论与数理统计简明教程. 北京: 高等教育出版社, 1988.

[9] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程. 北京: 高等教育出版社, 2004.

[10] 那汤松. 实变函数论. 第 5 版. 徐瑞云, 译. 北京: 高等教育出版社, 2010.

[11] 徐传胜. 从博弈问题到方法论学科——概率论发展史研究. 北京: 科学出版社, 2010.

[12] 陈善林, 张浙. 统计发展史. 上海: 立信会计图书用品社, 1987.

[13] 劳. 统计与真理——怎样运用偶然性. 北京: 科学出版社, 2004.

[14] 伊夫斯. 数学史概论, 第六版. 欧阳绛, 译. 哈尔滨: 哈尔滨工业大学出版社, 2009.

[15] 李文林. 数学珍宝——历史文献精选. 北京: 科学出版社, 2000.

[16] 塔巴克. 概率论和统计学——不确定性的科学. 杨静, 译. 北京: 商务印书馆, 2007.

[17] 淺野長一郎, 江島伸興, 李賢平. 基本統計學. 東京: 森北出版株式會社, 1993.

[18] 丸山儀四郎等. 確率および統計: 数學演習講座 10. 東京: 共立出版株式會社, 1958.

[19] 野中敏雄, 笹井敏夫. 確率・統計の演習. 東京: 森北出版株式會社, 1961.

[20] Uspensky J V. Introduction to Mathematical Probability. New York: McGraw-Hill, 1937.

[21] Rohatgi V K. An Introduction to Probability Theory and Mathematical Statistics. New York: John Wiley & Sons, 1976.

[22] von Plato J. Creating Modern Probability. Cambridge: Cambridge University Press, 1994.

[23] Kac M. Statistical Independence in Probability, Analysis and Number Theory. New York: John Wiley & Sons, 1959.

[24] Chow Y S. and Teicher H. Probability Theory. New York: Springer-Verlag, 1978.

[25] Billingsley P. Probability and Measure, second edition. New York: Wiley & Sons, 1986.

[26] Chung K L. A Course in Probability Theory, second edition. New York: Academic Press, 1974.

[27] Shao J. Mathematical Statistics, second edition. New York: Springer, 2003.

[28] Hald A. A History of Probability & Statistics and Their Applications Before 1750. New York: John Wiley and Sons, 1990.

[29] Fristedt B. and Gray L. A Modern Approach to Probability Theory. Boston: Birkhaüser, 1997.

[30] Gut A. Probability: A Graduate Course. New York: Springer, 2005.

[31] 梅斯特洛夫. 概率论史纲要. 张泽兴, 李贤平, 译 (未正式出版) .

[32] 李贤平. 投资分析. 复旦大学管理学院讲义, 2005.

[33] 李贤平. 金融工程. 复旦大学管理学院讲义, 2005.

［Ｇｅｎｅｒａｌ Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝概率论基础学习指导书
作者＝李贤平，陈子毅编
页数＝４２９
出版社＝北京市：高等教育出版社
出版日期＝２０１１．０7
ＳＳ号＝１２８８５０３７
ＤＸ号＝０００００８１４５０５３
ＵＲＬ＝ｈｔｔｐ：／／ｂｏｏｋ．ｓｚｄｎｅｔ．ｏｒｇ．ｃｎ／ｂｏｏｋＤｅｔａｉｌ．ｊｓｐ？ｄｘＮｕｍｂｅｒ＝００００
０８１４５０５３＆ｄ＝ＦＤ３２ＢＡＦ５２１Ｄ１Ｄ６１２４Ａ３０６Ａ５７ＤＦＥ１３ＥＥ０